PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/12, 9/00, C07K 14/47, C12N 15/52, C07K 16/18, 16/40, A61K 49/00, G01N 33/68

A1

(11) International Publication Number: WO 98/48010

(43) International Publication Date: 29 October 1998 (29.10.98)

(21) International Application Number: PCT/AU98/00280

(22) International Filing Date: 20 April 1998 (20.04.98)

(30) Priority Data:
PO 6334 21 April 1997 (21.04.97) AU

(81) Designated States: AU, CA, JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(71) Applicant (*for all designated States except US*): GARVAN INSTITUTE OF MEDICAL RESEARCH [AU/AU]; St. Vincent's Hospital, 384 Victoria Street, Darlinghurst, NSW 2010 (AU).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): CALLAGHAN, Michelle, J. [AU/AU]; 2/25-31 Alexandra Street, Drummoyne, NSW 2047 (AU). SUTHERLAND, Robert, L. [NZ/AU]; 20 Northcote Road, Lindfield, NSW 2070 (AU). WATTS, Colin, K., W. [NZ/AU]; 167 Avalon Parade, Avalon, NSW 2107 (US).

(74) Agent: F B RICE & CO.; 605 Darling Street, Balmain, NSW 2041 (AU).

(54) Title: NOVEL HUMAN TUMOUR SUPPRESSOR GENE

(57) Abstract

A novel human progestin-regulated gene designated EDD (E3 isolated by Differential Display) is disclosed which encodes a product exhibiting significant amino acid sequence identity with the HYD protein (*hyperplastic discs*) from *Drosophila melanogaster* and the 100 kDa HECT (homologous to E6-AP carboxyl terminus) domain protein from rat. The EDD gene appears to represent a tumour suppressor gene and the detection of a polymorphism or alteration in the gene from a subject may be useful for the diagnosis or determination of a predisposition to hyperproliferative disease such as a cancer. An assay for assessing progestin-responsiveness in a subject is also disclosed.

1	MTSIHFVVHP	LPGTEDQLND	RLREVSEKLN	KYNLNSHPPL	NVLEQATIKQ	CVVGPNHAAF
61	LLEDGRVCRI	GFSVQPDRLE	LGKPDNNDGS	KLNSNSGAGR	TSRPGRTSDS	PWFLSGSETL
121	GRLAGNTLGS	RWSSGVGGSG	GGSSGRSSAG	ARDSRRQTRV	IRTGRDRGSG	LLGSQPQPVI
181	PASVIPEELI	SQAQVVLQGK	SRSVIIRELQ	RTNLDVNLAV	NNLLSRDDED	GDDGDDTASE
241	SYLAGEDLMS	LLDADIHSAH	PSVIIDADAM	FSEDISYFGY	PSFRRSSLSR	LGSSRVLLLP
301	LERDSELLRE	RESVLRLRER	RWLDGASFDN	ERGSTSKEGE	PNLDKKNTPV	QSPVSLGEDL
361	QWWPDKDGTK	FICIGALYSE	LLAVSSKGEL	YQWKWSESEP	YRNAQNPSLH	HPRATFLGLT
421	NEKIVLLSAN	SIRATVATEN	NKVATWVDET	LSSVASKLEH	TAQTYSELQG	ERIVSLHCCA
481	LYTCAQLENS	LYWWGVVPFS	QRKKMLEKAR	AKNKKPKSSA	GISSMPNITV	GTQVCLRNNP
541	LYHAGAVAFS	ISAGIPKVGV	LMESVWNMND	SCRFQLRSPE	SLKNMEKASK	TTEAKPESKQ
601	EPVKTEMGPP	PSPASTCSDA	SSIASSASMP	YKRRRSTPAP	KEEEKVNEEQ	WSLREVVFVE
661	DVKNVPVGKV	LKVDGAYVAV	KFPGTSSNTN	CQNSSGPDAD	PSSLLQDCRL	LRIDELQVVK
721	TGGTPKVPDC	FQRTPKKLCI	PEKTEILAVN	VDSKGVHAVL	KTGNWVRYCI	FDLATGKAEQ
781	ENNFPTSSIA	FLGQNERNVA	IFTAGQESPI	ILRDGNGTIY	PMAKDCMGGI	RDPDWLDLPP
841	ISSLGMGVHS	LINLPANSTI	KKKAAVIIMA	VEKQTLMQHI	LRCDYEACRQ	YLMNLEQAVV
901	LEQNLQMLQT	FISHRCDGNR	NILHACVSVC	FPTSNKETKE	EEEAERSERN	TFAERLSAVE
961	AIANAISVVS	SNGPGNRAGS	SSSRSLRLRE	MMRRSLRAAG	LGRHEAGASS	SDHQDPVSPP
1021	IAPPSWVPDP	PAMDPDGDID	FILAPAVGSL	TTAATGTGQG	PSTSTIPGPS	TEPSVVESKD
1081	RKANAHFILK	LLCDSVVLQP	YLRELLSAKD	ARGMTPFMSA	VSGRAYPAAI	TILETAQKIA
1141	KAEISSSEKE	EDVFMGMVCP	SGTNPDDSPL	YVLCCNDTCS	FTWTGAEHIN	QDIFECRTCG
1201	LLESLCCCTE	CARVCHKGHD	CKLKRTSPTA	YCDCWEKCKC	KTLIAGQKSA	RLDLLYRLLT
1261	ATNLVTLPNS	RGEHLLLFLV	QTVARQTVEH	CQYRPPRIRE	DRNRKTASPE	DSDMPDHDLE
1321	PPRFAQLALE	RVLQDWNALK	SMIMFGSQEN	KDPLSASSRI	GHLLPEEQVY	LNQQSGTIRL
1381	DCFTHCLIVK	CTADILLLDT	LLGTLVKELQ	NKYTPGRREE	AIAVTMRFLR	SVARVFVILS
1441	VEMASSKKKN	NFIPQPIGKC	KRVFQALLPY	AVEELCNVAE	SLIVPVRMGI	ARPTAPFTLA
1501	STSIDAMQGS	EELFSVEPLP	PRPSSDQSSS	SSQSQSSYII	RNPQQRRISQ	SQPVRGRDEE
1561	QDDIVSADVE	EVEVVEGVAG	EEDHHDEQEE	HGEENAEAEG	QHDEHDEDGS	DMELDLLAAA
1621	ETESDSESNH	SNQDNASGRR	SVVTAATAGS	EAGASSVPAF	FSEDDSQSND	SSDSDSSSSQ
1681	SDDIEQETFM	LDEPLERTTN	SSHANGAAQA	PRSMQWAVRN	TQHQRAASTA	PSSTSTPAAS
1741	SAGLIYIDPS	NLRRSGTIST	SAAAAAAALE	ASNASSYLTS	ASSLARAYSI	VIRQISDLMG
1801	LIPKYNHLVY	SQIPAAVKLT	YQDAVNLQNY	VEEKLIPTWN	WMVSIMDSTE	AQLRYGSALA
1861	SAGDPGHPNH	PLHASQNSAR	RERMTAREEA	SLRTLEGRRR	ATLLSARQGM	MSARGDFLNY
1921	ALSLMRSHND	EHSDVLPVLD	VCSLKHVAYV	FQALIYWIKA	MNQQTTLDTP	QLERKRTREL
1981	LELGIDNEDS	EHENDDDTNQ	SATLNDKDDD	SLPAETGQNH	PFFRRSDSMT	FLGCIPPNPF
2041	EVPLAEAIPL	ADQPHLLQPN	ARKEDLFGRP	SQGLYSSSAS	SGKCLMEVTV	DRNCLEVLPT
2101	KMSYAANLKN	VMNMQNRQKK	EGEEQPVLPE	ETESSKPGPS	AHDLAAQLKS	SLLAEIGLTE
2161	SEGPPLTSFR	PQCSFMGMVI	SHDMLLGRWR	LSLELFGRVF	MEDVGAEPGS	ILTELGGFEV
2221	KESKFRREME	KLRNQQSRDL	SLEVDRDRDL	LIQQTMRQLN	NHFGRRCATT	PMAVHRVKVT
2281	FKDEPGEGSG	VARSFYTAIA	QAFLSNEKLP	NLECIQNANK	GTHTSLMQRL	RNRGERDRER
2341	EREREMRRSS	GLRAGSRRDR	DRDFRRQLSI	DTRPFRPASE	GNPSDDPEPL	PAHRQALGER
2401	*LYPRVQAMQP*	*AFASKITGML*	*LELSPAQLLL*	*LLASEDSLRA*	*RVDEAMELII*	AHGRENGADS
2461	ILDLGLVDSS	EKVQQENRKR	HGSSRSVVDM	DLDDTDDGDD	NAPLFYQPGK	RGFYTPRPGK
2521	NTEARLNCFR	NIGRILGLCL	LQNELCPITL	NRHVIKVLLG	RKVNWHDFAF	FDPVMYESLR
2581	QLILASQSSD	ADAVFSAMDL	AFAIDLCKEE	GGGQVELIPN	GVNIPVTPQN	VYEYVRKYAE
2641	HRMLVVAEQP	LHAMRKGLLD	VLPKNSLEDL	TAEDFRLLVN	GCGEVNVQML	ISFTSFNDES
2701	GENAEKLLQF	KRWFWSIVEK	MSMTERQDLV	YFWTSSPSLP	ASEEGFQPMP	SITIRPPDDQ
2761	HLPTANTCIS	RLYVPLYSSK	QILKQKLLLA	IKTKNFGFV		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece	**ML**	Mali	**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**MN**	Mongolia	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MR**	Mauritania	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MW**	Malawi	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MX**	Mexico	**US**	United States of America
CA	Canada	**IT**	Italy	**NE**	Niger	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NL**	Netherlands	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NO**	Norway	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NZ**	New Zealand	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**PL**	Poland		
CM	Cameroon	**KR**	Republic of Korea	**PT**	Portugal		
CN	China	**KZ**	Kazakstan	**RO**	Romania		
CU	Cuba	**LC**	Saint Lucia	**RU**	Russian Federation		
CZ	Czech Republic	**LI**	Liechtenstein	**SD**	Sudan		
DE	Germany	**LK**	Sri Lanka	**SE**	Sweden		
DK	Denmark	**LR**	Liberia	**SG**	Singapore		
EE	Estonia						

NOVEL HUMAN TUMOUR SUPPRESSOR GENE

Field of the Invention:

This invention relates to a novel human progestin-regulated gene designated EDD (E3 isolated by Differential Display) which encodes a product exhibiting significant amino acid sequence identity with the HYD protein (*hyperplastic discs*) from *Drosophila melanogaster* and the 100 kDa HECT (homologous to E6-AP carboxyl terminus) domain protein from rat.

Background to the Invention:

The control of cell proliferation and differentiation in the normal breast and in breast cancer involves complex actions and interactions of steroid hormones (in particular estrogen and progesterone), peptide hormones and growth factors (1, 2). How these agents act at critical control points within the cell cycle to influence progression through the cycle or exit to enter a pathway of differentiation is only partially understood (3-5).

Progestins are responsible for mammary gland lobuloalveolar development during pregnancy (6), although there is evidence for a more predominant role for estrogens than progestins in stimulating epithelial cell proliferation in the normal premenopausal breast (7, 8). Progestins both stimulate and inhibit breast cancer epithelial cell proliferation *in vitro* but the predominant effect is growth inhibition probably via induction of differentiation (3, 4, 7, 9). Progestin action is mediated primarily through the progesterone receptor (PR), which acts as a transcriptional transactivator for a largely undefined set of progestin-responsive genes which may, in turn, transcriptionally or post-transcriptionally influence additional genes or gene products.

Only a limited number of genes have been implicated in progestin action on cell proliferation. Previous studies by the present inventors have identified c-*myc* and cyclin D1 as major downstream targets of progestin-stimulated cell cycle progression in human breast cancer cells (3, 10) while the delayed growth inhibitory effects of progestins involve decreases in cyclin D1 and E gene expression (4, 9). While progestin effects on c-*myc* gene expression are rapid and occur within minutes, effects on cyclin expression begin several hours later, pointing to the presence of undefined earlier events.

Since progestin action is complex and is likely to involve multiple genes, many of which are currently unknown, the differential display RT-PCR technique (DD-PCR) (11) was adopted to identify target genes in cultured human breast cancer cells. The utility of this approach has been previously demonstrated by the cloning of PRG1, a gene having significant homology with isoforms of 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase (12). Using the same technique, a novel progestin-regulated gene, EDD (designated DD5 in the applicant's Australian Provisonal patent application No. PO6334), has been identified.

Based on amino acid sequence similarity, EDD appears to be a human homologue of the *Drosophila* tumor suppressor gene *hyperplastic discs* (*hyd*) (13). Although the function of the HYD protein is unknown, significant homology exists between its carboxyl terminus and those of human E6-AP and a number of proteins identified through database searches (14). These HECT domain family proteins function as ubiquitin-protein ligases (E3 enzymes) (14-16), playing a role in the ubiquitination cascade that targets specific substrate proteins for proteolysis. Notably, the protein encoded by EDD has a carboxy-terminal HECT domain containing a cysteine residue that covalently binds ubiquitin. This amino acid is conserved in all known HECT domain-containing E3 enzymes and is involved in the transfer of ubiquitin. It is therefore proposed that the EDD gene represents a novel human tumour suppressor gene encoding a ubiquitin-protein ligase.

Disclosure of the Invention:

In a first aspect, the present invention provides an isolated polynucleotide molecule comprising a nucleotide sequence encoding a protein which comprises the following N-terminal amino acid sequence:

MTSIHFVVHP

or a biologically active portion of said protein.

Preferably, the encoded protein comprises the following N-terminal amino acid sequence:

MTSIHFVVHPLPGTEDQLNDRLREVSEKLNKYNLNSHPPLNVLEQATIKQ.

More preferably, the encoded protein is ubiquitin-protein ligase and has an approximate molecular weight of 300kDa.

Most preferably, the isolated polynucleotide molecule comprises a nucleotide sequence substantially corresponding to or, at least, ≥ 90% (more

preferably, ≥ 95%) homologous to the nucleotide sequence shown at Figure 3B from nucleotide 34 to nucleotide 8424 or a portion(s) thereof.

The term "portion(s) thereof" in this regard is to be understood as referring to portion(s) of the nucleotide sequence which encode biologically active peptide or polypeptide portions or antigenic determinants. Typically, such "portions(s) thereof" will comprise a nucleotide sequence of at least 50 nucleotides in length. However, shorter portions of the nucleotide sequence (e.g. portions of ≥ 8 nucleotides in length) may also be used in or for the production of probes useful for hybridization assays.

Thus, in a second aspect, the present invention provides an oligonucleotide or polynucleotide probe molecule labelled with a suitably detectable label (e.g. radioisotopes), comprising a nucleotide sequence substantially corresponding to, or complementary to, a ≥ 8 nucleotide portion of the nucleotide sequence shown at Figure 3B from nucleotide 34 to nucleotide 8424.

Such probe molecules may be DNA or RNA. They may be used, for example, to quantitatively or qualitatively detect EDD mRNA in total or poly(A) RNA isolated from one or more tissues. As discussed below, such assays may have diagnostic and/or prognostic value.

The present invention also further extends to oligonucleotide primers for the above sequences, antisense sequences and homologues of said primers and antisense sequences, complementary ribozyme sequences, catalytic antibody binding sites and dominant negative mutants of the polynucleotide molecules.

Preferably, the polynucleotide molecule of the first aspect is of human origin. More preferably, the polynucleotide molecule is of human cancer cell origin.

The isolated polynucleotide molecule of the first aspect may be incorporated into plasmids or expression vectors or cassettes, which may then be introduced into suitable bacterial, yeast, insect or mammalian host cells. Such host cells may be used to express the protein or biologically active fragment thereof encoded by the isolated polynucleotide molecule.

As mentioned above, the amino acid sequence of the EDD product (pEDD) shows significant sequence similarity to the amino acid sequence of the HYD protein of *Drosophila*. The *Drosophila hyd* gene is a tumour suppressor gene and it is therefore expected that the EDD gene is similarly a

tumour suppressor gene. Further, it is expected that the pEDD protein will have activity similar to the HYD protein. Particularly, inactivating or other mutations in EDD may give rise to susceptibility to cancer, thus making EDD a potential target for preventive or therapeutic strategies. Mutations in EDD could also be diagnostic for cancer susceptibility, particularly for early diagnosis in normal or pre-neoplastic disease or be useful in predicting tumour progression or response to therapy (i.e. a prognostic marker). Further, since EDD is likely to be involved in cell cycle regulation by progestins and other mitogens, EDD is a potential target for antiproliferative agents (i.e. cancer therapeutics). Moreover, as EDD is one of only a few known genes to be regulated by progestins, EDD is an important mediator of progestin action and a marker of clinical responsiveness to progestins.

As a tumour suppressor gene, EDD could be a familial cancer susceptibility gene, for example, like p16 (Multiple Tumor Suppressor Gene 1, MTS1) or the familial breast cancer susceptibility gene BRCA1. It might also have a role in sporadic cancer.

In a third aspect, the present invention provides in a substantially pure form, a protein (designated pEDD) comprising the following N-terminal amino acid sequence:

MTSIHFVVHP

or a biologically active portion of said protein.

Preferably, the protein of the third aspect comprises the following N-terminal amino acid sequence:

MTSIHFVVHPLPGTEDQLNDRLREVSEKLNKYNLNSHPPLNVLEQATIKQ.

More preferably, the protein of the third aspect is a ubiquitin-protein ligase and has an approximate molecular weight of 300kDa.

Most preferably, the protein of the third aspect comprises an amino acid sequence substantially corresponding to the amino acid sequence shown in Figure 3C.

The biologically active portions may consist of polypeptide or peptide sequences which inhibit, mimic or enhance the biological effect of the protein. Additionally, the biologically active portions may also represent antigenic determinants useful for raising antibodies specific to the protein.

The protein, or biologically active portion thereof, according to the third aspect may be purified from natural sources (e.g. whole brain, heart,

testis and appendix) or suitable cell lines, or may be produced recombinantly by any of the methods common in the art (Sambrook *et al.*, 1989).

In a fourth aspect, the present invention provides a non-human organism transformed with the polynucleotide molecule of the first aspect of the present invention.

The organisms which may be usefully transformed with the polynucleotide molecule of the first aspect include bacteria such as *E. coli* and *B. subtilis*, eukaryotic cell lines such as CHO, fungi and plants.

In a fifth aspect, the present invention provides an antibody specific to the protein designated pEDD or an antigenic portion thereof.

The antibody may be polyclonal or monoclonal and may be produced by any of the methods common in the art.

It is also to be understood that the invention relates to kits for diagnostic assays, said kits comprising a protein or biologically active portion thereof according to the second aspect and/or an antibody according to the fifth aspect. Additionally, or alternatively, the kit may comprise oligonucleotide probes for hybridisation assays or oligonucleotide primers for PCR based assays.

In a sixth aspect, the present invention provides a protein or antigenic portion thereof, capable of binding to an anti-pEDD antibody.

As will be seen hereinafter, in some tissues EDD appears to be regulated by progestin. EDD may, therefore, provide a useful marker for progestin-responsiveness in a subject. For example, as a marker of breast or endometrial tumour or meningioma responsiveness to progestins or progestin antagonists (antiprogestins) - i.e. high levels may indicate that the tumour is responsive to progestins/antiprogestins and could be sensitive to progestin/antiprogestin therapy. EDD may also be a useful prognostic marker since hormonally responsive tumours often have a better prognosis (i.e. patients have longer disease-free survival and overall survival). Alternatively, mutations, deletions or amplification of the EDD gene might predict tumour progression, and disease prognosis independent of its role a progestin-regulated gene. Thus, levels of EDD mRNA present in isolated cells or tissue samples may be assessed by DNA or RNA probes or primers in hybridisation assays or PCR analysis. Alternatively, the level of pEDD protein may be assessed through the use of the abovementioned antibodies.

Thus, in a seventh aspect, the present invention provides as assay for assessing progestin-responsiveness in a subject comprising the steps of;

(i) isolating cells or tissue from said subject; and

(ii) detecting the presence of a protein comprising an amino acid sequence substantially corresponding to that shown at Figure 3C .

In some circumstances, it may be preferred to expose the isolated cells or tissue to progestin or agonist or antagonist compounds and, subsequently, determine whether the progestin or agonist or antagonist compound has induced the production of the pEDD protein.

In an eighth aspect, the present invention provides a method for the diagnosis or determination of a predisposition to hyperproliferative disease, especially cancer, comprising detecting in a subject a polymorphism or alteration in the EDD gene which is indicative of said hyperproliferative disease or a predisposition to said hyperproliferative disease or developmental abnormality.

The modulation of EDD activity may also have therapeutic utility in the treatment of proliferative disorders, such as malignant or non-malignant hyperproliferative disease (e.g. breast and other cancers), and dermatological diseases or developmental abnormalities. Further, modulation of EDD may be of therapeutic value in processes involving progestin action in progestin target organs (e.g. fertility control, and reproductive tissue function).

EDD activity could be regulated by:

- synthetic compounds, either stimulatory or inhibitory (i.e. agonists or antagonists),

- ribozymes specific for EDD (i.e. to down-regulate endogenous EDD activity), and

- gene therapy using expression vectors or oligonucleotides or other delivery systems (e.g. viral) containing a nucleotide sequence coding for EDD sense (i.e. to augment endogenous pEDD protein levels and activity) or antisense (i.e. to down-regulate endogenous pEDD protein levels and activity). Sense vectors could contain only a portion of the EDD coding sequence if separate desirable activities are found to reside in separate portions of the protein. Such vectors could also include dominant negative mutants of EDD which encode a gene product causing an altered phenotype by, for example, reducing or eliminating the activity of the endogenous pEDD protein. This might be caused through the interuption of formation of

enzyme complexes, substrate competition or the formation of a defective substrate or reaction product. Particular examples of dominant negative mutants may be mutants that encode truncated proteins retaining pEDD sequences involved in protein-protein interactions or substrate recognition but which lack enzymatic or other activities residing elsewhere in the pEDD protein. Expression of such mutants would inhibit correct substrate modification or processing. Thus as a putative ubiquitin-protein ligase, truncated pEDD proteins could be expressed which allow the binding of protein substrates but which lack the sequences necessary for the subsequent ubiquitination and destruction of these sequences.

Since the pEDD protein seems likely to be involved in cell cycle (growth) regulation including cell proliferation, differentiation and cell death, the pEDD protein or an agonist or antagonist might be used as a chemoprotectant in cancer chemotherapy treatments. That is, the pEDD protein or agonist/antagonist may be administered to a patient so as to stop the cell cycle including cell proliferation, differentiation and cell death in normal cells prior to treatment with standard cancer drugs (e.g. methotrexate, vinblastine and cisplatin). The arrested cells would thereby be less prone to damage by chemotherapy toxicity.

The term "substantially corresponding" as used herein in relation to the nucleotide sequence is intended to encompass minor variation(s) in the nucleotide sequence which due to degeneracy in the DNA code do not result in a change in the encoded protein. Further, this term is intended to encompass other minor variations in the sequence which may be required to enhance expression in a particular system but in which the variation(s) do not result in a decrease in biological activity of the encoded protein.

The term "substantially corresponding" as used herein in relation to amino acid sequence is intended to encompass minor variations in the amino acid sequence which do not result in a decrease in biological activity of the encoded protein. These variation(s) may include conservative amino acid substitution(s). The substitution(s) envisaged are:-

G,A,V,I,L,M; D, E; N,Q; S,T; K,R,H; F,Y,W,H; and P,Nα-alkalamino acids.

The terms "comprise", "comprises" and "comprising" as used throughout the specification, are intended to refer to the inclusion of a stated step, component or feature or group of steps, components or features with or

without the inclusion of a further component or feature or group of steps, components or features.

The invention will hereinafter be further described by way of the following non-limiting example and accompanying figures.

__Brief description of the accompanying figures__:

Figure 1. Identification of a differentially expressed cDNA in T-47D cells treated with the synthetic progestin ORG 2058.

A) Identification of EDD by differential display. Total RNA obtained from T-47D cells treated with ORG 2058 or vehicle control (ethanol) for 3 h was used as a template for differential display PCR reactions. The PCR products were separated on a 4.5% polyacrylamide denaturing gel and visualized by autoradiography. The arrow indicates the EDD DD-PCR product (DD5-1; see Fig. 3A) which is present at a higher level in the progestin treated (ORG 2058) compared with control lane.

B) Confirmation of the progestin induction of EDD by Northern blot analysis. T-47D cells proliferating in medium supplemented with 5% charcoal-treated FCS were treated with 10 nM ORG 2058 or ethanol vehicle (CONTROL) in the presence or absence of actinomycin D (ACT) and after 3 h total RNA was harvested for Northern analysis. The Northern blot was probed with the EDD clone P19.

C) Effect of cycloheximide on progestin induction of EDD mRNA. T-47D cells proliferating in medium supplemented with 5% charcoal-treated FCS were treated with ORG 2058 (10 nM), cycloheximide (CHX, 20 μg/ml), ORG 2058 and CHX simultaneously or ethanol vehicle and harvested for total RNA at 1 h. The Northern blot was probed with the EDD DD-PCR fragment DD5-1.

Figure 2. Expression of EDD mRNA in human tissues.

A) Northern blot analysis of polyA$^+$ RNA from human tissues. The blot was hybridized with the P19 cDNA clone of EDD. Molecular sizes of markers are indicated. PBL, peripheral blood leukocytes.

B) Dot blot analysis of polyA$^+$ RNA from human tissues. The blot was hybridized with the P19 cDNA clone of EDD. Row A: 1, whole brain; 2, amygdala; 3, caudate nucleus; 4, cerebellum; 5, cerebral cortex; 6, frontal lobe; 7, hippocampus; 8, medulla oblongata; Row B: 1, occipital lobe; 2, putamen; 3, substantia nigra; 4, temporal lobe; 5, thalamus; 6, sub-thalamic

nucleus; 7, spinal cord; Row C: 1, heart; 2, aorta; 3, skeletal muscle; 4, colon, 5, bladder; 6, uterus; 7, prostate; 8, stomach; Row D: 1, testis; 2, ovary; 3, pancreas; 4, pituitary gland; 5, adrenal gland; 6, thyroid gland; 7, salivary gland; 8, mammary gland; Row E: 1, kidney; 2, liver; 3, small intestine; 4, spleen; 5, thymus; 6, peripheral leukocyte; 7, lymph node; 8, bone marrow; Row F: 1, appendix; 2, lung; 3, trachea; 4, placenta; Row G: 1, fetal brain; 2, fetal heart; 3, fetal kidney; 4, fetal liver; 5, fetal spleen; 6, fetal thymus; 7, fetal lung.

Figure 3. Cloning and predicted amino acid sequence of EDD.

A) A schematic representation of EDD structure with a restriction map for the EDD cDNA indicating the sites used for cloning the full-length EDD construct and the cDNA clones used to derive the EDD sequence shown beneath. The DD-PCR cDNA fragment identified by differential display was designated DD5-1 and a cDNA clone derived from the 5' RACE product and the original DD-PCR product, DD5-2. All cDNA clones were isolated from a human placenta cDNA library with the exception of H1 which was isolated from a human heart cDNA library.

B) The nucleotide sequence of EDD. The start and stop codons are underlined.

C) Predicted amino acid sequence of pEDD. There are two regions with high homology (~60%) to HYD (a central sequence and a carboxyl sequence containing the HECT domain) and these and other highly conserved sequences are shown in bold type, while two putative nuclear localization signals are boxed. The HECT domain is underlined and in bold type and includes a conserved cysteine at residue 2768 (boxed). A region showing homology to polyA-binding proteins is italicized and the peptide sequence to which antiserum AbPEP1 was raised is underlined. The numbers refer to positions of amino acids.

Figure 4. Chromosomal localization of the EDD gene.

Metaphase showing FISH with the H1 probe. Normal male chromosomes were stained with DAPI. Hybridization sites on chromosome 8 are indicated by an arrow.

Figure 5. Characterization of EDD protein.

A) Detection of recombinant EDD protein with AbPEP1. *Sf*9 cells infected with baculovirus containing a truncated EDD construct (EDD 100 kDa) were boiled in SDS-sample buffer prior to SDS-PAGE through a 6% gel,

transferred to nitrocellulose and blotted with AbPEP1 or AbPEP1 peptide-blocked.

B) Determination of the size of the EDD protein. EDD was immunoprecipitated from T-47D lysate using AbPEP1. The immunoprecipitate (IP) was resolved by SDS-PAGE through a 6% gel alongside the products of *in vitro* translated full length EDD (IVT) and immunoprecipitated *in vitro* translated EDD (IVT-IP). The T-47D immunoprecipitate was transferred to nitrocellulose and blotted for EDD with AbPEP1 while the remainder of the gel was dried and autoradiographed. Molecular masses of marker proteins are indicated.

C) Detection of EDD protein in T-47D lysates. Immunoprecipitated EDD was run alongside 40 μg total protein from T-47D lysate. Total proteins were blotted with either AbPEP1 or peptide-blocked AbPEP1 and the immunoprecipitate was blotted with AbPEP1.

Figure 6. EDD protein expression in human tissues and cell lines.

Expression of EDD in normal breast and breast cancer cell lines. Total cell lysates from a range of cell lines were separated by SDS-PAGE through a 6% gel, transferred to nitrocellulose and blotted with AbPEP1. 184 is a normal breast cell line, 184B5 an immortalized derivative, and the remainder are breast cancer cell lines, MCF-7M being a sub-line of MCF-7.

Figure 7. Sequence of the rat 100 kDa protein cDNA.

Autoradiograph of the sequencing gel obtained when one clone was sequenced using the EDD-specific FC2 primer, with the sequence (a) listed alongside the autoradiograph. The published sequence (b,) is shown alongside and the missing base denoted by an asterisk.

Figure 8. Ubiquitin thiol ester formation by EDD.

In vitro translation of truncated (A) or full-length (B) EDD wild type or mutant (C2768A) protein in the presence of ^{35}S-methionine was followed by a 10 min incubation at 25 °C either with or without purified GST-ubiquitin (or GST in part A) fusion protein. Samples were resolved by SDS-PAGE (A, 7% gel; B, 6% gel) following either incubation at 25 °C for 20 min in non-reducing sample buffer containing 4 M urea or boiling in sample buffer containing 100 mM DTT. Ubiquitin- and GST-ubiquitin-bound forms are marked with arrows.

Example:

MATERIALS AND METHODS

Reagents

Steroids and growth factors were obtained from the following sources: ORG 2058 (16a-ethyl-21-hydroxy-19-norpregn-4-en-3,20-dione), Amersham Australia Pty Ltd, Sydney, Australia; human transferrin, Sigma Chemical Co., St. Louis, Mo.; and human insulin, Actrapid, CSL-Novo, North Rocks, Australia. Steroids were stored at -20 °C as 1000-fold-concentrated stock solutions in absolute ethanol. Cycloheximide (Calbiochem-Behring Corp., La Jolla, CA) was dissolved at 20 mg/ml in water and filter sterilized. Actinomycin D (Cosmegen, Merck Sharp and Dohme Research Pharmaceuticals, Rahway, NJ) was dissolved at 0.5 mg/ml in sterile water and used immediately. Tissue culture reagents were purchased from standard sources.

Cell culture

The sources and maintenance of the human breast cancer and normal cell lines used were as described previously (12, 22), as were tissue culture experiments (12). Briefly, progestin (ORG 2058, 10 nM) and/or cycloheximide (20 μg/ml) or actinomycin D (5 μg/ml) was added to the medium and control flasks received the same volume of vehicle alone. To obtain RNA for differential display, cells were grown in insulin-supplemented serum-free medium and treated for 3 h with ORG 2058 or ethanol vehicle. Subsequent progestin stimulation experiments were carried out in medium containing 5% charcoal-stripped fetal calf serum without insulin.

RNA isolation and Northern analysis:

Cells harvested from duplicate 150 cm^2 flasks were pooled, RNA extracted by a guanidinium-isothiocyanate-cesium chloride procedure and Northern analysis was performed as previously described with 20 μg of total RNA per lane (3, 23). The membranes were hybridized overnight (50 °C) with probes labelled with [a-^{32}P]dCTP (Amersham Australia Pty Ltd) using a Prime-a-Gene labelling kit (Promega Corp., Sydney, Australia). The membranes were washed at a highest stringency of 0.2 ¥ SSC (30 mM NaCl, 3 mM sodium citrate [pH 7.0]) / 1% sodium dodecyl sulfate at 65 °C and exposed to Kodak X-OMAT or BIOMAX film at -70 °C. Human multiple tissue Northern blots or RNA Master blot (CLONTECH Laboratories Inc., Palo Alto, CA) were hybridized under conditions recommended by the manufacturer. The mRNA abundance was quantitated by densitometric analysis of autoradiographs using Molecular

Dynamics Densitometer and software (Molecular Dynamics, Sunnyvale, CA). The accuracy of loading was estimated by re-hybridizing membranes with a [g-^{32}P]ATP end-labelled oligonucleotide complementary to 18S rRNA (24, 25).

Differential display

Differential display was carried out as previously described (11) using a Heiroglyph mRNA Profile Kit No. 1 (Genomyx Corporation, Foster City, CA) and recommended protocol. First strand cDNA synthesis was carried out in 96-well format 0.2 ml thin walled tubes. Typically 200 ng total RNA from T-47D cells treated with the synthetic progestin ORG 2058 for 3 h or from control T-47D cells was reverse transcribed with Expand Reverse Transcriptase enzyme (Boehringer Mannheim Pty Ltd, Castle Hill, Australia) following annealing with 4 pmol anchored primer (5'ACGACTCACTATAGGGCT$_{12}$AC). Subsequent PCR amplification was performed with one-tenth of the resultant cDNA in duplicate reactions containing [a-^{33}P] dATP with the anchored primer (0.2 μM), an arbitrary primer (5'ACAATTTCACACAGGAGCTAGCAGAC, 0.2 μM) and Expand Long Template Taq DNA Polymerase (Boehringer Mannheim). The PCR products were denatured and separated on a 4.5% denaturing polyacrylamide gel at 800 v for 16 h using the Genomyx Long Read Sequencing System reagents and apparatus. The gel was dried on the glass plate and exposed to X-ray film for 16-72 h. The DD-PCR product of interest was excised from the gel and amplified by PCR under the conditions recommended by the kit manufacturer using an M13 forward primer (5'AGCGGATAACAATTTCACACAGGA) and a T7 promoter primer (5'TAATACGACTCACTATAGGG). The reamplified PCR products were purified from 0.8% agarose gels using QIAEX reagents (Qiagen Pty Ltd, Clifton Hill, Australia).

Cloning and sequencing of cDNAs

Double stranded DNA templates were sequenced using the *fmol* DNA Cycle Sequencing System (Promega Corp.) with [^{33}P]-labelled primers. The M13 primer was used for direct sequencing of DD-PCR products and the T7 and SP6 (5'GATTTAGGTGACACTATAG) promoter primers were used for sequencing PCR products cloned into the pGEM-T vector (Promega Corp.). Sequence database searches were performed at the NCBI using the Blast or Fasta network services. Peptide motif searches were carried out against the Prosite database.

Two primers (FC2: 5'GACGAAGGGCCCTGACTGCGCGAGAAGAAGC and R2: 5'AAAGAATTCTGTCATGGAGTCTGAACGTCG) that flank the region

containing the reported rat 100 kDa start codon (26) were used to amplify cDNA extracted from a rat hypothalamus library (CLONTECH). The resulting PCR product was cloned into pGEM-T (Promega Corp.) and four clones were sequenced.

Rapid amplification of cDNA 5' ends (5'RACE)

Additional sequence was obtained with the aid of a 5'RACE kit (Life Technologies Inc., Gaithersburg, MD), following the manufacturer's instructions. Briefly, a gene specific primer (GSP1: 5'CACGCTCCAATGCAAGCTGG) was used to prime first strand cDNA synthesis. Following removal of the RNA strand, cDNA was 5' poly dC tailed and amplified by PCR. The target cDNA was amplified using an anchor primer (UAP: 5'GGCCACGCGTCGACTAGTACGGGIIGGGIIGGGIIG, where I represents deoxyinosine) in combination with a second gene specific primer (GSP2: 5'CGATCTTCCCTGATTCGAGGTGGC). Various gel-purified PCR products were further PCR amplified, primed by UAP and a third gene specific nested primer (GSP3: 5'CTGTATTGACAATGCTCCACC).

cDNA library screening

10^6 plaques from a human heart cDNA library in the Lambda ZAPII vector primed with both oligo (dT) and random primers (Stratagene, La Jolla, CA) were transferred to nylon membranes (Hybond N, Amersham Australia Pty Ltd) and screened with both the original DD-PCR fragment and the RACE product as [^{32}P]-labelled probes. This led to isolation of clone H1 (2.55 kb). This clone and the RACE product were used to screen 10^6 recombinants from a human placenta 5'-STRETCH PLUS cDNA library in lgt10 primed with both oligo (dT) and random primers (CLONTECH Laboratories, Inc.). Sequencing of cDNA clones in either pBluescript or lgt10 was carried out as described above using vector-specific or gene-specific primers. Several rounds of isolation of positive clones and further screening of this library led to the isolation of the following overlapping clones covering the entire EDD open reading frame: P61 (1.95 kb), P43 (2.1 kb), P1 (1.5 kb), P19 (3 kb) and P47 (2.1 kb).

Fluorescence *in situ* hybridization

A probe corresponding to clone H1 was nick-translated with biotin-14-dATP and hybridized *in situ* at a final concentration of 20 ng/ml to metaphases from two normal males. The fluorescence *in situ* hybridization (FISH) method was modified from that previously described (27) in that chromosomes were stained before analysis with both propidium iodide (as counterstain) and DAPI

(for chromosome identification). Images of metaphase preparations were captured by a CCD camera using the CytoVision Ultra image collection and enhancement system (Applied Imaging Int Ltd). FISH signals and the DAPI banding pattern were merged for figure preparation.

Construction of recombinant cDNA clones for *in vitro* translation and protein expression

The full length EDD sequence was cloned by ligating three PCR products which spanned the open reading frame into pBluescript. The existing *Sal*I and *Eco*RI restriction sites used to ligate the fragments are indicated in Fig. 3A. The carboxyl third of the cDNA was cloned into pBluescript such that an 890 amino acid truncated protein corresponding to the predicted rat 100 kDa protein (from aa 1910 to aa 2799) would be translated. An identical truncated cDNA fragment was cloned into the pFASTBAC 1 expression vector (Life Technologies Inc.) for protein expression using the BAC-TO-BAC baculovirus expression system in *Spodoptera frugiperda* (*Sf*9) cells and full length EDD cDNA was cloned into the pRcCMV expression vector (Invitrogen, Leek, The Netherlands) for transient transfection into HEK-293 cells. Mutagenesis of cysteine 2768 to alanine was performed for full length and truncated constructs in pBluescript using the Quick-Change site-directed mutagenesis kit (Stratagene). *In vitro* transcription and translation were performed using the TNT T7 Quick coupled rabbit reticulocyte lysate system (Promega Corp.) and [^{35}S]-methionine (1000 Ci/mmole, ICN Biomedicals Australasia Pty Ltd, Seven Hills, Australia).

SDS-polyacrylamide gel electrophoresis (PAGE) and immunoblotting

Cells growing in mid-log phase were lysed in 1% Triton X100 buffer containing 50 mM 4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES; pH 7.5), 150 mM NaCl, 10% glycerol, 1.5 mM $MgCl_2$, 1 mM EGTA, 10 mM sodium pyrophosphate, 20 mM sodium fluoride, 1 mM dithiothreitol (DTT), 10 μg/ml each of aprotonin and leupeptin, 1 mM phenylmethylsulfonyl fluoride (PMSF) and 200 μM sodium orthovanadate. Lysates were cleared by centrifugation, quantitated according to a modified Bradford method (Bio-Rad Laboratories, Hercules, CA) and typically 40 μg of total protein in SDS-sample buffer (50 mM Tris-HCl, pH 6.8, 2% SDS, 10% glycerol, and 0.2% bromophenol blue) containing 5% b-mercaptoethanol were resolved on 6% SDS-polyacrylamide gels. Following electrophoresis proteins were transferred to nitrocellulose (TransBlot, Bio-Rad Laboratories) and subjected to immunodetection. An EDD-specific peptide (SSEKVQQENRKRHGSS) was

synthesised, coupled via glutaraldehyde to diptheria toxoid and used to generate a rabbit anti-EDD antibody (designated AbPEP1).

Immunoprecipitation

Cleared cell lysates (typically 1 mg total protein) or *in vitro* translation reactions were incubated with either control rabbit serum or AbPEP1 in the presence or absence of a 10-fold excess of competing peptide for 1-2 hr at 4 °C. Following incubation with Protein A Sepharose 4B (Zymed, San Francisco, CA), immunoprecipitates were washed three times in 1% Triton X100 lysis buffer described above, resolved by SDS-PAGE and either transferred to nitrocellulose and immunoblotted with AbPEP1 or where applicable dried onto Whatman 3MM paper and subjected to autoradiography.

Ubiquitin-binding assay

[^{35}S]-labelled *in vitro* translated truncated (~100 kDa) or full length protein was tested for its ability to bind ubiquitin by incubating 5 μl translation reaction with or without 5 μg purified GST protein or GST-ubiquitin fusion protein for 10 min at 25 °C (28). Reactions were terminated by incubating the mixtures in either SDS-sample buffer containing 100 mM DTT at 95 °C for 5 min or in SDS-sample buffer containing 4 M urea instead of DTT at 25 °C for 20 min. Samples were resolved by SDS-PAGE through 6% or 7% gels followed by drying and autoradiography.

RESULTS

Isolation and Northern blot analysis of a progestin regulated cDNA

The differential display technique was used to identify mRNAs in T-47D human breast cancer cells with altered levels of expression in response to treatment with the synthetic progestin ORG 2058 for 3 h. When the anchored primer, 5'ACGACTCACTATAGGGCT$_{12}$AC was used in conjunction with the arbitrary primer, 5'ACAATTTCACACAGGAGCTAGCAGAC, a cDNA fragment of approximately 850 bp that was more abundant in treated samples than in control samples was identified and designated EDD (Fig. 1A). Northern analysis of total cellular RNA from T-47D cells showed that transcription was required for the observed ORG 2058 induction of EDD mRNA levels as this was blocked in the presence of actinomycin D (Fig. 1B). Induction was also prevented by cycloheximide, suggesting that EDD is not directly transcriptionally regulated by progestin acting via the PR (Fig. 1C).

The tissue specificity of EDD gene expression was investigated by hybridizing Northern blots of polyA+ RNA isolated from human tissues to the EDD cDNA fragment. A single transcript of 9.5 kb was detected in a variety of tissues (Fig. 2A) with the highest expression in testis, heart, placenta and skeletal muscle. Hybridization to a more quantitatively loaded RNA dot blot (Fig. 2B) confirmed that EDD is expressed at varying levels in all tissues examined and that the mRNA was most abundant in testis and expressed at high levels in brain, pituitary and kidney. Significant levels of expression were also observed in placenta, uterus, prostate, stomach, fetal lung and various brain tissues. EDD mRNA was also expressed in a range of breast cancer cell lines, not all of which are progestin-responsive (not shown).

Cloning of the full length EDD cDNA

The original DD5-1 fragment isolated by DD PCR was 850 bp in length and is shown schematically in Figure 3A. The DNA sequence of this fragment had no homology to sequences of any known human genes. To obtain the complete coding sequence from which EDD was derived a combination of 5'RACE and screening of human heart and placenta cDNA libraries was used. This resulted in a series of overlapping clones covering 8.5 kb of sequence (Fig. 3A; Genbank Accession AF006010). Analysis of the nucleotide sequence (Fig. 3B) revealed an open reading frame of 2799 amino acids (Fig. 3C). The EDD sequence was divided into overlapping 1800 bp segments and used in Blastx searches of the GenBank database. The only homology to a human sequence of known function was to polyA binding protein across 50 amino acids (50%, Fig. 3C) although the similarities among mammalian polyA binding proteins in this stretch are usually in the vicinity of 100%.

The DNA sequence of EDD showed significant similarity to two sequences in the database. Both of these genes encode proteins belonging to the HECT family of ubiquitin-protein ligases, although their specificities are unknown. HECT proteins contain a conserved domain of approximately 300 amino acids that contains a cysteine residue able to bind ubiquitin via a thioester linkage. Nucleotides 5667 to 8502 of EDD were 88% identical to the rat 100 kDa protein cDNA sequence (26), nucleotides 572 to 740 and 3498 to 3867 were 69% identical to two regions of the *Drosophila melanogaster hyperplastic discs* gene (*hyd*) and nucleotides 7560 to 8430 were 60% identical to a third region of *hyd* (13). The putative initiation codon is

surrounded by a consensus sequence for strong translational initiation (ACCATGA, (29)) and corresponds to a possible start codon of the Drosophila *hyd* gene (13). The stop codon corresponds to that shared by the rat 100 kDa protein and *hyd* genes. Like EDD, both the *hyd* and rat 100 kDa protein genes have estimated mRNA transcript sizes of 9.5 kb (14, 26). The predicted EDD protein is identical to HYD at 40% of amino acid residues and similar at 64% of residues, while the carboxyl third of EDD is 96% identical and 98.5% similar to rat 100 kDa protein. The most highly conserved regions between HYD and EDD are designated by bold type in Figure 3C. Within two of these regions there are stretches of 40-80 amino acids that are highly conserved between HYD, EDD and a possible *C. elegans* homologue of HYD contained within 2 overlapping cosmids (Genbank Accession No. G1729554 and G1729549). The longest conserved regions between EDD and HYD are a central domain of approximately 400 amino acids (58% identity, 72% similarity) and the carboxyl 300 amino acids which include the HECT domain and conserved cysteine residue (64% identity, 80% similarity). This latter region also showed around 30% identity and 50% similarity with other HECT proteins including yeast RSP5 or PUB-1 and RAD26 (14, 30, 31), and the mammalian proteins UreB1 (19), *Nedd*-4 (15, 20, 32, 33) and E6-AP (15, 17, 18). Apart from two putative nuclear localization signals (34), no other consensus functional domains were identified within the EDD sequence.

Chromosomal localization of the EDD gene

FISH was used to localize the gene for EDD. Eighteen metaphases from a normal male were examined for fluorescent signal. Seventeen of these metaphases showed signal on one or both chromatids of chromosome 8 in the region q22. High resolution studies of 8 metaphases showed signal at q22.3 (Fig. 4). There was a total of 4 non-specific background dots observed in these 18 metaphases. A similar result was obtained from hybridization of the probe to 11 metaphases from a second normal male (data not shown). This localization was consistent with independent assignment of an EST corresponding to EDD (EST116344) using the radiation hybrid panel Genebridge 4.

Characterization of EDD protein

A rabbit antiserum (AbPEP1) against an EDD-specific peptide matching a sequence towards the carboxyl terminus of the protein (underlined in Figure 3C) reacted strongly on Western blots with a truncated

(100 kDa) recombinant EDD protein expressed in *Sf9* cells using a baculovirus system (Fig. 5A). A second strongly reactive band of approximately 200 kDa was also seen, but this appeared to be non-specific as antibody binding was not competed by the EDD peptide. The full length EDD cDNA was cloned into pBluescript and translated *in vitro* in a rabbit reticulocyte lysate system. The size of the major product was in agreement with the expected molecular mass of the protein as predicted from the amino acid sequence (~300 kDa, Fig. 5B). The identity of the translated protein was confirmed by immunoprecipitation from either translation reactions or T-47D whole cell lysates with AbPEP1 (Fig. 5B). Western blotting of whole cell lysates from T-47D cells using AbPEP1 detected two major bands, both abolished in the presence of competing peptide - a major species at approximately 230 kDa and a minor species of higher molecular mass (Fig. 5C). This latter band corresponds in size to that of the *in vitro* translated protein and is immunoprecipitated by AbPEP1 (Fig. 5C) and by two other EDD-specific peptide antibodies (not shown). However, the 230 kDa protein is not immunoprecipitated from cell lysates by these antibodies. As a single EDD mRNA transcript was detected on Northern blots, it was hypothesised that the EDD protein may be processed to the 230 kDa form which could be folded in such a way that was not susceptible to immunoprecipitation in its native state. However, transient expression of full length EDD in HEK-293 cells followed by Western blotting of whole cell lysates revealed an increase in the expression of the 300 kDa species only (not shown). Western blotting of whole cell lysates from a number of normal breast and breast cancer epithelial cell lines showed that EDD protein was expressed in all immortalized and cancer cell lines but not in a normal breast cell line, 184 (Fig. 6).

Identity of the rat gene product

The previously described rat cDNA that is highly homologous to the EDD gene reportedly gives rise to a 100 kDa protein, inferred from cDNA sequence data which showed several in-frame stop codons upstream of the putative initiation codon (26), corresponding to amino acid residue 1910 of EDD. These stop codons were not present in the EDD cDNA. Furthermore, although we were able to confirm that an anti-HYD antibody detected an approximately 100 kDa protein in rat muscle lysates, this species was not detected by AbPEP1 even though the predicted sequences of human and rat

proteins are identical at every residue of the peptide used to raise the AbPEP1 antibody. This led the present inventors to question whether the 100 kDa protein was the actual rat gene product.

A segment of rat cDNA was cloned containing the stretch of sequence upstream of the proposed initiation codon and found an additional base that, by changing the reading frame, removes the upstream stop codons (Fig. 7). Correction of this apparent error results in a rat cDNA sequence that closely matches the human cDNA, in which a continuous open reading frame exists throughout the sequence. While the rat cDNA sequence corresponding to the amino terminal two-thirds of EDD has not been cloned, a number of mouse expressed sequences covering parts of this region are recorded in the GenBank database (Accession No. AA183561, AA177260, AA183970, AA231351, AA087561) and these show similar levels of similarity with the EDD DNA sequence as that seen with the published rat sequence. Thus it appears that the true product of the rat gene is not a 100 kDa protein but may exist as a larger species. In rat lysates, however, AbPEP1 does not detect a protein having a molecular weight consistent with the human (EDD) and *Drosophila* (HYD) gene products.

Ubiquitin binding by EDD

A critical feature of the HECT family of E3 enzymes is their ability to reversibly form thioesters with ubiquitin at a conserved cysteine residue within the HECT domain. This property has been demonstrated for the HECT proteins human E6-AP, rat 100 kDa protein and yeast RSP5 where the thioester linkage remains intact in the absence of reducing agents but is broken in the presence of 100 mM DTT (14). Substitution of the conserved cysteine residue prevents ubiquitin thioester bond formation. However, this property has not been shown for the HYD protein. To assess the potential of EDD to function as an E3 we tested whether EDD could form a reversible bond with ubiquitin via the conserved cysteine, C2768. ^{35}S-labelled *in vitro* translated truncated protein (~100 kDa of carboxyl terminus sequence) was incubated with purified GST-ubiquitin fusion protein in the presence or absence of DTT before SDS-PAGE (Fig. 8A).

In the absence of DTT an additional higher molecular mass protein band was observed that corresponded to the expected size of an EDD-GST-ubiquitin conjugate (~ 130 kDa, upper arrow in Fig. 8A). This species was abolished in the presence of 100 mM DTT suggesting involvement of a

thioester bond in its formation. This was confirmed by experiments with an *in vitro* translated protein containing a C2768A mutation: binding of GST-ubiquitin was not seen under these conditions (Fig. 8A). A species of slightly higher molecular mass than EDD was also observed (lower arrow in Fig. 8A), consistent with the formation of ubiquitin-EDD conjugates, ubiquitin being present as a component of the rabbit reticulocyte lysate. Again this was not observed using the mutant protein or in the presence of 100 mM DTT. Similar results were achieved with full length EDD protein obtained (though at lower yield) by *in vitro* translation (Fig. 8B).

DISCUSSION

Application of the differential display PCR technique to a cultured human breast cancer cell model in which clearly defined proliferative responses to progestins are observed has led to the identification of a novel gene, EDD, a likely human homologue of the *Drosophila melanogaster* tumor suppressor gene *hyperplastic discs* (13). EDD is also highly homologous to the partial published sequence for the cDNA encoding the rat 100 kDa protein (26). All three genes produce large (approx 9.5 kb) mRNAs and the predicted entire EDD open reading frame of 2799 amino acids shares 40% identity with that of *Drosophila hyd* while the carboxyl-terminal 889 amino acids of EDD share 96% identity with the rat protein. Western analysis showed that the EDD gene product is a protein of approximately 300 kDa. This protein is also immunoprecipitated by 3 different peptide-specific EDD antibodies and also corresponds to the size of the major *in vitro* translated gene product. The large discrepancy in the predicted size of the human and rat proteins was apparently resolved by re-examination of the rat cDNA sequence which disclosed an error in the published translation start site, pointing to the likelihood that a larger gene product exists.

At their carboxyl termini EDD, its rat homologue and HYD all contain a highly homologous HECT domain, indicating membership of a larger family of proteins which function as ubiquitin protein ligases (E3s). The ubiquitination of target proteins occurs by the action of multiple interacting proteins: a ubiquitin-activating enzyme (E1), ubiquitin-conjugating enzymes (E2) and ubiquitin-protein ligases (E3). Substrate specificity is largely determined by E3s, which bind and transfer ubiquitin to the target protein following interaction with specific E2s. The key feature of the HECT class

of E3s is their ability to covalently bind ubiquitin through a conserved cysteine residue located in their HECT domains (14). This property was demonstrated for EDD using *in vitro* translated protein that lost the ability to bind ubiquitin if the conserved cysteine (C2768) was substituted and it was therefore concluded that EDD is an E3.

Few E3 genes have been cloned (only two from human) but others are likely to exist as ubiquitin-dependent proteolysis is involved in many cellular processes and targets many known proteins. Ubiquitin-mediated proteolysis is critical in the control of cell cycle progression, being responsible for the periodic destruction of key cell cycle regulators including cyclins (35-37) and cyclin-dependent kinase inhibitors (38, 39) and also targeting transcription factors (40-43), the tumor suppressor protein p53 (18) and cell-cell signalling components such as b-catenin (44). Disruption of the murine Itch locus, which encodes an E3, caused hyperplasia in lymphoid and gastrointestinal epithelial tissues and an abnormal inflammatory response (21) while mutations in E6-AP in humans result in neurological abnormalities, indicating critical, and perhaps tissue specific, roles for E3 proteins (45).

Although substrates for EDD and its rat and *Drosophila* homologs have yet to be defined, conservation between the central domain of EDD and that of HYD suggests that this region has an important function, perhaps in substrate recognition. For the yeast E3 Rsp5, substrate specificity is determined by the amino terminal domain and does not require the HECT domain (16). Alternatively, this region could be involved in the binding of as yet unknown E2 proteins that interact specifically with EDD. The mouse E3 *Nedd4* has at least two distinct E2 binding domains, only one of which is within the HECT domain (15) while human E6-AP requires only the HECT domain for E2 recognition (46). As the protein produced from the truncated EDD construct still binds ubiquitin reversibly, at least some E2 recognition function is present in this carboxyl domain. Other possible functions of the conserved central domain include cellular localization or translocation between cytoplasm and nucleus, cofactor association or phosphorylation.

Although ubiquitination is clearly involved in steroid-responsive processes such as regulation of cell cycle progression, specific regulation of ubiquitin pathways by steroid hormones has not previously been reported. The precise role of EDD in progestin action is unknown, particularly whether

it participates in those key early events that occur in response to this hormone and which are ultimately responsible for its effects on cell proliferation and differentiation. Progestin regulation of EDD mRNA, which requires *de novo* protein synthesis, is transient with maximal levels 3 to 4-fold above control observed at 6 h. This increase in EDD expression levels therefore precedes the increase in the S phase fraction of T-47D cells following ORG 2058 treatment under the same conditions, which typically occurs at 12 to 14 h (3) and hence is consistent with a possible role in control of cell cycle progression. Similar levels of EDD induction were observed in antiestrogen-arrested MCF-7 breast cancer cells treated with 17b-estradiol (not shown), suggesting this may be a generalized response to mitogens.

However, given that EDD is also expressed in non-progestin target tissues, a more widespread role than specifically mediating progestin effects is expected. Information on the biological role of HYD from mutagenesis studies in *Drosophila* (13) may ultimately give clues as to the function of EDD. The null *hyd* phenotype is lethal, as are severe mutations in the pupal or larval stages. Less severe mutations result in overgrowth (hyperplasia) of larval imaginal discs (the larval centres of cell proliferation that give rise to adult structures such as wings, legs and antennae), apparently caused by a failure to terminate cell proliferation when the discs reach their characteristic size, hence the definition of *hyd* as a tumor suppressor gene. Surviving adults are sterile due to germ cell defects, and interestingly, high expression levels of EDD and rat 100 kDa protein mRNA are seen in human and rat testes, suggesting a critical function in this organ.

Studies of a number of human homologues of *Drosophila* tumor suppressor genes strongly suggests that these genes have similar roles in both species in controlling cell proliferation, and that such genes can be important in human heritable and sporadic cancers, for example *patched* (47), mutations of which are linked to basal cell carcinoma, and *discs large* (45, 48), a target of the APC gene which is mutated in sporadic colorectal tumors and familial adenomatous polyposis coli. The possible involvement of EDD in human tumorigenesis and tumor progression is therefore of particular interest. The EDD gene locus at chromosome 8q22 is often disrupted in a variety of cancers, being deleted in adenocarcinoma of the ovary and lung (49, 50), hepatocellular carcinoma (51) and head and neck squamous cell carcinoma (52), amplified in many tumor types including gastrointestinal and

primary breast cancers (53, 54) and involved in translocations in acute myeloid leukemia (55). Chromosome 8q22 is also a region affected in the human developmental disorder Klippel-Feil syndrome (56).

REFERENCES

1. Clarke, C.L., and Sutherland, R.L. (1990) *Endocr. Rev.* 11, 266-301

2. Dickson, R.B., and Lippman, M.E. (1995) *Endocr Rev* 16, 559-589

3. Musgrove, E.A., Lee, C.S.L., and Sutherland, R.L. (1991) *Mol Cell Biol* 11, 5032-5043

4. Musgrove, E.A., Swarbrick, A., Lee, C.S.L., Cornish, A.L., and Sutherland, R.L. (1998) *Mol. Cell. Biol.* In Press,

5. Prall, O.W.J., Sarcevic, B., Musgrove, E.A., Watts, C.K.W., and Sutherland, R.L. (1997) *J. Biol. Chem.* 272, 10882-10894

6. Lydon, J.P., DeMayo, F.J., Funk, C.R., Mani, S.K., Hughes, A.R., Montgomery Jr., C.A., Shyamala, G., Conneely, O.M., and O'Malley, B.W. (1995) *Genes Dev.* 9, 2266-2278

7. Sutherland, R.L., Prall, O.W.J., Watts, C.K.W., and Musgrove, E.A. (1998) *J. Mamm. Gland Biol. Neoplasia* 3, 63-71

8. Laidlaw, I.J., Clarke, R.B., Howell, A., Owen, A.W.M.C., Potten, C.S., and Anderson, E. (1995) *Endocrinology* 136, 164-171

9. Groshong, S.D., Owen, G.I., Grimison, B., Schauer, I.E., Todd, M.C., Langan, T.A., Sclafani, R.A., Lange, C.A., and Horwitz, K.B. (1997) *Mol. Endocrinol.* 11, 1593-1607.

10. Musgrove, E.A., Hamilton, J.A., Lee, C.S.L., Sweeney, K.J.E., Watts, C.K.W., and Sutherland, R.L. (1993) *Mol. Cell. Biol.* 13, 3577-3587

11. Liang, P., and Pardee, A.B. (1992) *Science* 257, 967-971

12. Hamilton, J.A., Callaghan, M.J., Sutherland, R.L., and Watts, C.K.W. (1997) *Mol. Endocrinol.* 11, 490-502

13. Mansfield, E., Hersperger, E., Biggs, J., and Shearn, A. (1994) *Dev. Biol.* 165, 507-526

14. Huibregtse, J.M., Scheffner, M., Beaudenon, S., and Howley, P.M. (1995) *Proc Natl Acad Sci U S A* 92, 5249

15. Hatakeyama, S., Jensen, J.P., and Weissman, A.M. (1997) *J Biol Chem* 272, 15085-15092

16. Huibregtse, J.M., Yang, J.C., and Beaudenon, S.L. (1997) *Proc. Natl. Acad. Sci. U S A* 94, 3656-3661

17. Huibregtse, J.M., Scheffner, M., and Howley, P.M. (1993) *Mol. Cell. Biol.* 13, 775-784

18. Scheffner, M., Huibregtse, J.M., Vierstra, R.D., and Howley, P.M. (1993) *Cell* 75, 495-505

19. Gu, J., Ren, K., Dubner, R., and Iadarola, M.J. (1994) *Mol Brain Res* 24, 77-88

20. Kumar, S., Harvey, K.F., Kinoshita, M., Copeland, N.G., Noda, M., and Jenkins, N.A. (1997) *Genomics* 40, 435-443

21. Perry, W.L., Hustad, C.M., Swing, D.A., O'Sullivan, T.N., Jenkins, N.A., and Copeland, N.G. (1998) *Nat. Genet.* 18, 143-146

22. Buckley, M.F., Sweeney, K.J.E., Hamilton, J.A., Sini, R.L., Manning, D.L., Nicholson, R.I., deFazio, A., Watts, C.K.W., Musgrove, E.A., and Sutherland, R.L. (1993) *Oncogene* 8, 2127-2133

23. Alexander, I.E., Clarke, C.L., Shine, J., and Sutherland, R.L. (1989) *Mol Endocrinol* 3, 1377-1386

24. Chan, Y.-L., Guttell, R., Noller, H.F., and Wool, I.G. (1984) *J Biol Chem* 259, 224-230

25. Hall, R.E., Lee, C.S.L., Alexander, I.E., Shine, J., Clarke, C.L., and Sutherland, R.L. (1990) *Int J Cancer* 46, 1081-1087

26. Muller, D., Rehbein, M., Baumeister, H., and Richter, D. (1992) *Nucleic Acids Res* 20, 1471-1475

27. Callen, D., Baker, E., Eyre, H.J., Chernos, J.E., Bell, J.A., and Sutherland, G.R. (1990) *Ann Genet* 33, 219-221

28. Scheffner, M., Nuber, U., and Huibregtse, J.M. (1995) *Nature* 373, 81-83

29. Kozak, M. (1987) *Nucleic Acids Res* 15, 8125-8132

30. Nefsky, B., and Beach, D. (1996) *EMBO J.* 15, 1301-1312

31. van Gool, A.J., Verhage, R., Swagemakers, S.M., van de Putte, P., Brouwer, J., Troelstra, C., Bootsma, D., and Hoeijmakers, J.H. (1994) *EMBO J.* 13, 5361-5369

32. Staub, O., Dho, S., Henry, P., Correa, J., Ishikawa, T., McGlade, J., and Rotin, D. (1996) *EMBO J.* 15, 2371-2380

33. Nagase, T., Miyajima, N., Tanaka, A., Sazuka, T., Seki, N., Sato, S., Tabata, S., Ishikawa, K., Kawarabayasi, Y., Kotani, H., et al. (1995) *DNA Res* 2, 37-43

34. Dingwall, C., and Laskey, R.A. (1991) *Trends Biochem Sci* 16, 478-481

35. Won, K.A., and Reed, S.I. (1996) *EMBO J.* 15, 4182-4193

36. Diehl, J.A., Zindy, F., and Sherr, C.J. (1997) *Genes Dev* 11, 957-972

37. King, R.W., Deshaies, R.J., Peters, J.M., and Kirschner, M.W. (1996) *Science* 274, 1652-1659

38. Benito, J., Martin-Castellanos, C., and Moreno, S. (1998) *EMBO J.* 17, 482-497

39. Pagano, M., Tam, S.W., Theodoras, A.M., Beer-Romero, P., Del Sal, G., Chau, V., Yew, P.R., Draetta, G.F., and Rolfe, M. (1995) *Science* 269, 682-685

40. Musti, A.M., Treier, M., Peverali, F.A., and Bohmann, D. (1996) *Biol. Chem.* 377, 619-624

41. Musti, A.M., Treier, M., and Bohmann, D. (1997) *Science* 275, 400-402

42. Palombella, V.J., Rando, O.J., Goldberg, A.L., and Maniatis, T. (1994) *Cell* 78, 773-785

43. Orian, A., Whiteside, S., Israel, A., Stancovski, I., Schwartz, A.L., and Ciechanover, A. (1995) *J. Biol. Chem.* 270, 21707-21714

44. Orford, K., Crockett, C., Jensen, J.P., Weissman, A.M., and Byers, S.W. (1997) *J. Biol. Chem.* 272, 24735-24738

45. Kishino, T., Lalande, M., and Wagstaff, J. (1997) *Nat Genet* 15, 70-73

46. Kumar, S., Kao, W.H., and Howley, P.M. (1997) *J. Biol. Chem.* 272, 13548-13554

47. Johnson, R.L., Rothman, A.L., Xie, J., Goodrich, L.V., Bare, J.W., Bonifas, J.M., Quinn, A.G., Myers, R.M., Cox, D.R., Epstein, E., Jr., and Scott, M.P. (1996) *Science* 272, 1668-1671

48. Matsuura, T., Sutcliffe, J.S., Fang, P., Galjaard, R.J., Jiang, Y.H., Benton, C.S., Rommens, J.M., and Beaudet, A.L. (1997) *Nat Genet* 15, 74-77

49. Mitelman, F., Mertens, F., and Johansson, B. (1997) *Nat. Genet.* 15, 417-474

50. Sato, S., Nakamura, Y., and Tsuchiya, E. (1994) *Cancer Res.* 54, 5652-5655

51. Piao, Z., Park, C., Park, J., and Kim, H. (1998) *Int. J. Cancer* 75, 29-33

52. Nawroz, H., van der Riet, P., Hruban, R.H., Koch, W., Ruppert, J.M., and Sidransky, D. (1994) *Cancer Res.* 54, 11152-11155

53. El-Rifai, W., Sarlomo-Rikala, M., Miettinen, M., Knuutila, S., and Andersson, L.C. (1996) *Cancer Res* 56, 3230-3233

54. Muleris, M., Almeida, A., Gerbault-Seureau, M., Malfoy, B., and Dutrillaux, B. (1994) *Genes Chromosomes Cancer* 10, 160-170

55. Erickson, P., Gao, J., Chang, K.S., Look, T., Whisenant, E., Raimondi, S., Lasher, R., Trujillo, J., Rowley, J., and Drabkin, H. (1992) *Blood* 80, 1825-1831

56. Clarke, R.A., Singh, S., McKenzie, H., Kearsley, J.H., and Yip, M.Y. (1995) *Am J Hum Genet* 57, 1364-1370

The abbreviations used in this specification are: DD-PCR, differential display polymerase chain reaction; DTT, dithiothreitol; EDD, E3 isolated by differential display; FISH, fluorescence *in situ* hybridization; GST, glutathione S-transferase; HECT, homologous to E6-AP carboxyl terminus; PAGE, polyacrylamide gel electrophoresis; PMSF, phenylmethylsulfonyl fluoride; PR, progesterone receptor; RACE, rapid amplification of cDNA ends.

It will be appreciated by persons skilled in the art that numerous variations and/or modifications may be made to the invention as shown in the specific embodiments without departing from the spirit or scope of the invention as broadly described. The present embodiments are, therefore, to be considered in all respects as illustrative and not restrictive.

<u>**The Claims**</u>:

1. An isolated polynucleotide molecule comprising a nucleotide sequence encoding a protein which comprises the following N-terminal amino acid sequence:

   MTSIHFVVHP

   or a biologically active portion of said protein.

2. A polynucleotide molecule according to claim 1, wherein the encoded protein comprises the following N-terminal amino acid sequence: MTSIHFVVHPLPGTEDQLNDRLREVSEKLNKYNLNSHPPLNVLEQATIKQ.

3. A polynucleotide molecule according to claim 1 or 2, wherein the encoded protein is a ubiquitin-protein ligase and has an approximate molecular weight of 300kDa.

4. A polynucleotide molecule according to any one of claims 1 to 3, comprising a nucleotide sequence ≥ 90% homologous to the nucleotide sequence shown at Figure 3B from nucleotide 34 to nucleotide 8424 or a portion(s) thereof.

5. A polynucleotide molecule according to any one of claims 1 to 3, comprising a nucleotide sequence ≥ 95% homologous to the nucleotide sequence shown at Figure 3B from nucleotide 34 to nucleotide 8424 or a portion(s) thereof.

6. A polynucleotide molecule according to any one of claims 1 to 3, comprising a nucleotide sequence substantially corresponding to the nucleotide sequence shown at Figure 3B from nucleotide 34 to nucleotide 8424 or a portion(s) thereof.

7. An oligonucleotide or polynucleotide probe molecule labelled with a suitably detectable label, said probe molecule comprising a nucleotide sequence substantially corresponding to, or complementary to, a ≥ 8

nucleotide portion of the nucleotide sequence shown at Figure 3B from nucleotide 34 to nucleotide 8424.

8. An expression vector or cassette, said vector or cassette comprising a polynucleotide molecule according to any one of claims 1 to 6 operably linked to a promoter sequence.

9. A non-human organism, said organism stably transformed with a polynucleotide molecule according to any one of claims 1 to 6.

10. A non-human organism, said organism stably transformed with a expression vector or cassette according to claim 8.

11. A protein comprising the following N-terminal amino acid sequence:

MTSIHFVVHP

or a biologically active portion of said protein, said protein or biologically active portion thereof being in a substantially pure form.

12. A protein according to claim 11, wherein said protein comprises the following N-terminal amino acid sequence:
MTSIHFVVHPLPGTEDQLNDRLREVSEKLNKYNLNSHPPLNVLEQATIK
Q.

13. A protein according to claim 11 or 12, wherein said protein is a ubiquitin-protein ligase and has an approximate molecular weight of 300kDa.

14. A protein according to any one of claims 11 to 13, wherein the protein comprises an amino acid sequence substantially corresponding to the amino acid sequence shown in Figure 3C.

15. An antibody or fragment thereof which specifically binds to a protein according to any one of claims 11 to 14 or an antigenic portion thereof.

16. A protein or antigenic portion thereof capable of binding to an anti-pEDD antibody.

17. An assay for assessing progestin-responsiveness in a subject, said method comprising the steps of;
    (i) isolating cells or tissue from said subject; and
    (ii) detecting the presence of a protein comprising an amino acid sequence substantially corresponding to that shown at Figure 3C .

18. An assay according to claim 17, wherein before step (ii) the isolated cells or tissue is exposed to progestin or a progestin agonist or antagonist.

19. An assay according to claim 16 or 17, wherein said step (ii) is conducted using an antibody or fragment thereof according to claim 15.

20. A method for the diagnosis or determination of a predisposition to hyperproliferative disease, said method comprising detecting in a subject a polymorphism or alteration in a gene comprising a nucleotide sequence substantially corresponding to the nucleotide sequence shown in Figure 3B from nucleotide 34 to nucleotide 8424, said polymorphism or alteration being indicative of said hyperproliferative disease or a predisposition to said hyperproliferative disease.

21. A method according to claim 20, wherein said hyperproliferative disease is a cancer.

22. A method according to claim 21, wherein said cancer is breast cancer.

FIGURE 1 A

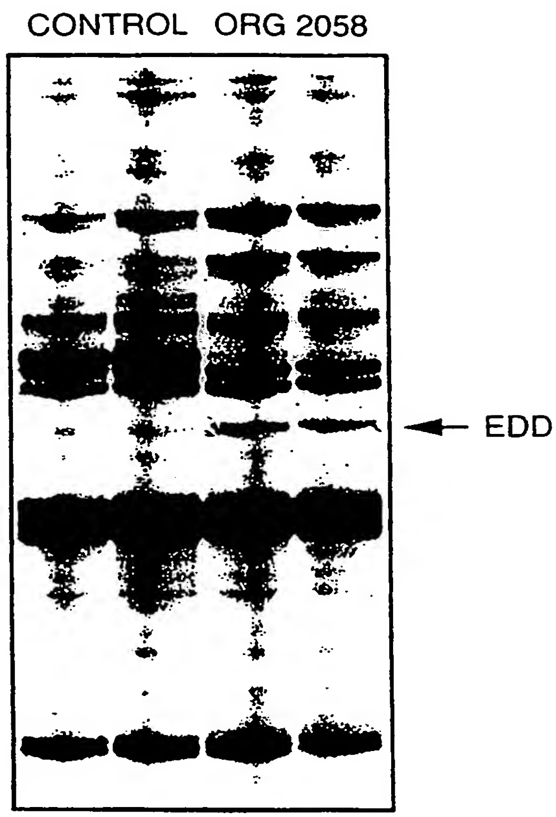


FIGURE 1B

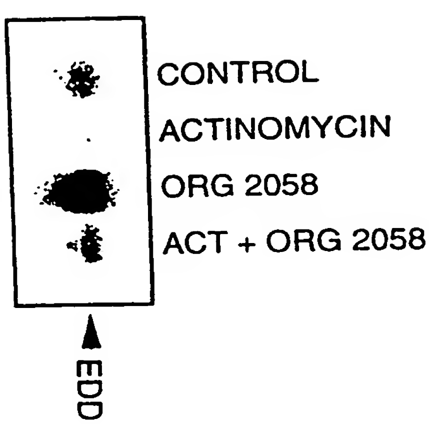
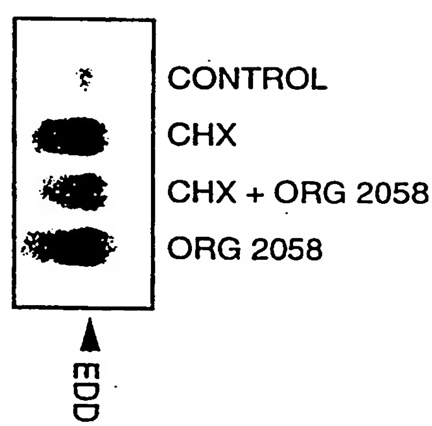


FIGURE 1C



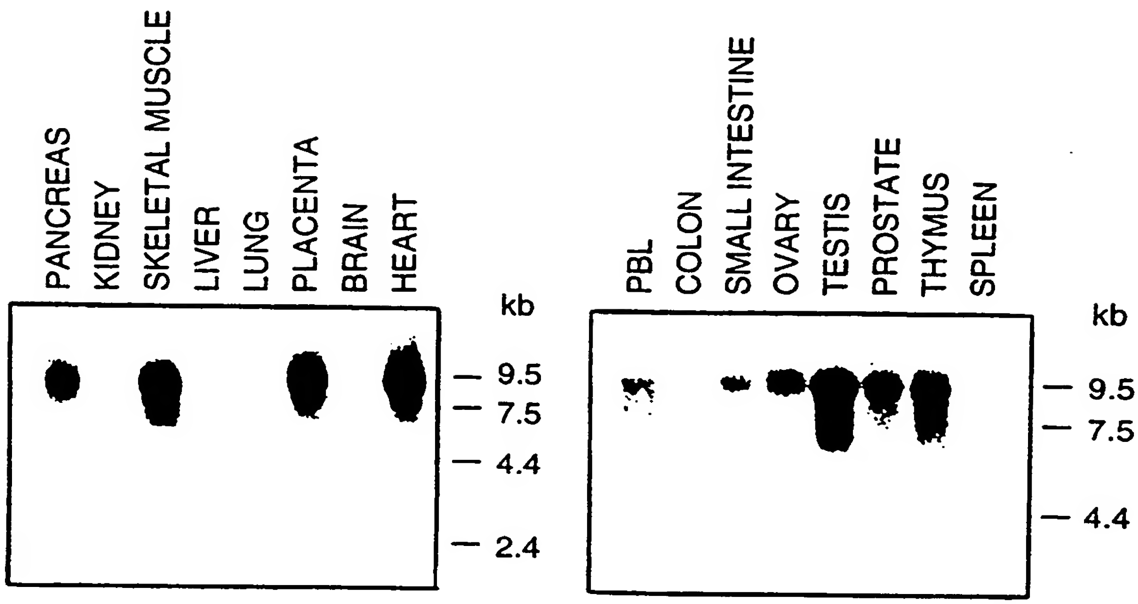
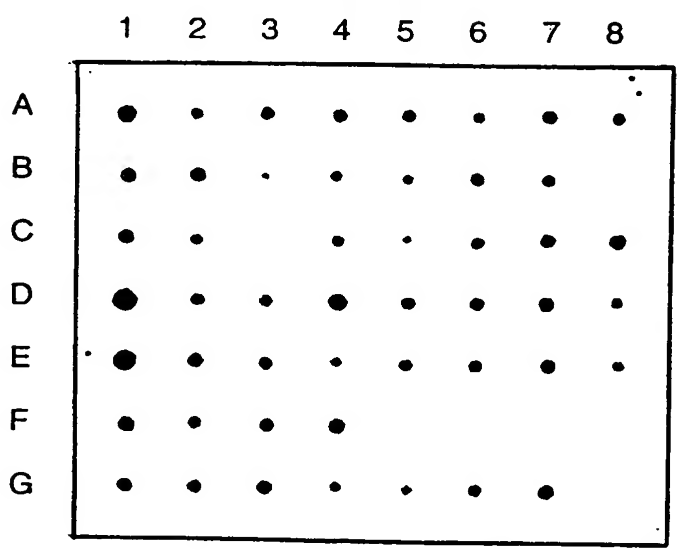


FIGURE 2A

FIGURE 2B



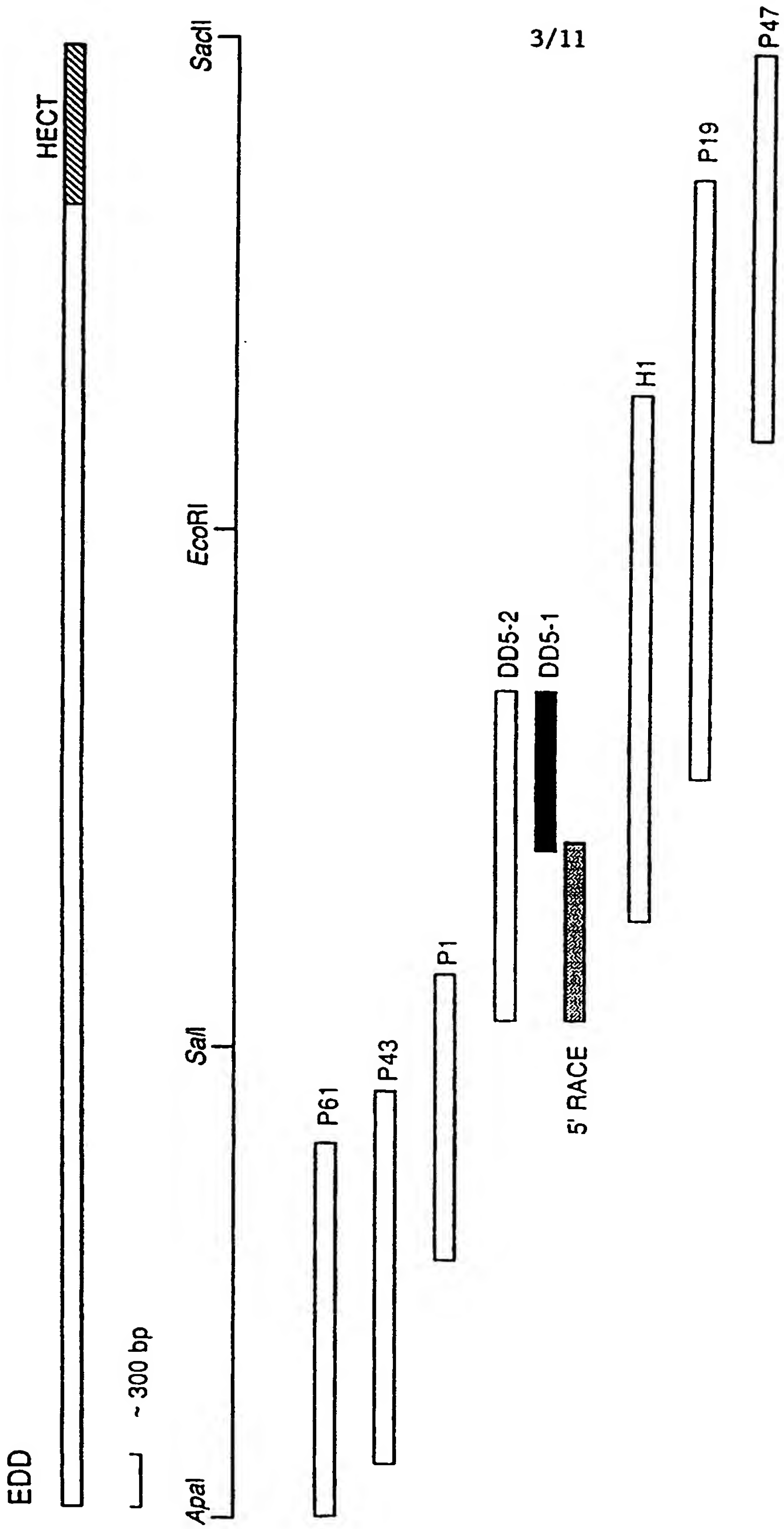

EDD
HECT
~ 300 bp
ApaI
SalI
EcoRI
SacII
P61
P43
P1
DD5-2
DD5-1
5' RACE
H1
P19
P47

FIGURE 3A

FIGURE 3B

```
   1 CGCCCTCGAG TGGAGGACGA GAAGGAAAGC ACCATGACGT CCATCCATTT CGTGGTTCAC
  61 CCGCTGCCGG GCACCGAGGA CCAGCTCAAT GACAGGTTAC GAGAAGTTTC TGAGAAGCTG
 121 AACAAATATA ATTTAAACAG CCACCCCCCT TTGAATGTAT TCCAACAGGC TACTATTAAA
 181 CAGTGTGTGG TGGGACCAAA TCATGCTGCC TTTCTTCTTG AGGATGGTAG AGTTTGCAGG
 241 ATTGGTTTTT CAGTACAGCC AGACAGATTG GAATTGGGTA AACCTGATAA TAATGATGGG
 301 TCAAAGTTGA ACAGCAACTC GGGGGCAGGG AGGACGTCAA GGCCTGGTAG GACAAGCGAC
 361 TCTCCATGGT TTCTCTCAGG TTCTGAGACT CTAGGCAGGC TGGCAGGCAA CACCTTAGGA
 421 AGCCGCTGGA GTTCTGGAGT GGGTGGAAGT GGTGGAGGAT CCTCTGGTAG GTCATCAGCT
 481 GGAGCTCGAG ATTCCCGCCG GCAGACTCGA GTTATTCGGA CAGGACGGGA TCGAGGGTCT
 541 GGGCTTTTGG GCAGTCAGCC CCAGCCAGTT ATTCCAGCAT CTGTCATTCC AGAGGAGCTG
 601 ATTTCACAGG CCCAAGTTGT TTTACAAGGC AAATCCAGAA GTGTCATTAT TCGAGAACTT
 661 CAGAGAACAA ATCTTGATGT GAACCTTGCT GTAAATAATT TACTTAGCCG GGATGATGAA
 721 GATGGAGATG ATGGGGATGA TACAGCCAGC GAATCTTATT TGGCTGGAGA GGATCTTATG
 781 TCTCTCCTTG ATGCCGACAT TCATTCTGCC CACCCAAGTG TCATTATTGA TGCAGATGCC
 841 ATGTTTTCTG AAGACATTAG CTATTTTGGT TACCCTTCTT TTCGTCGTTC ATCACTTTCC
 901 AGGCTAGGCT CATCTCGAGT TCTCCTTCTT CCCTTAGAGA GAGACTCTGA GCTGTTGCGT
 961 GAACGCGAAT CCGTTTTACG TTTACGTGAA CGAAGGTGGC TTGATGGAGC CTCATTTGAT
1021 AATGAAAGGG GTTCTACCAA GCAAGGAAGG AGAGCCAAAC TTGATAAGAA GAATACACCT
1081 GTTCAAAGTC CAGTATCTCT AGGAGAAGAT TTGCAGTGGT GGCCTGATAA GGATGGAACA
1141 AAATTCATCT GTATGGCTCT GTATTCTGAA CTTCTGGCTG TCAGCAGTAA AGGAGAACTT
1201 TATCAGTGGA AATGGAGTGA ATCTGAGCCT TACAGAAATG CCCAGAATCC TTCATTACAT
1261 CATCCACGAG CAACATTTTT GGGGTTAACC AATGAAAAGA TAGTCCTCCT GTCTGCAAAT
1321 AGCATAAGAG CAACTGTAGC TACAGAAAAG AACAAGGTTG CTACATGGGT GGATGAAACT
1381 TTAAGTTCTG TGGCTTCTAA ATTAGAGCAC ACTGCTCAGA CTTACTCTGA ACTTCAAGGA
1441 GAGCGGATAG TTTCTTTACA TTGCTGTGCC CTTTACACCT GCGCTCAGCT GGAAAACAGT
1501 TTATATTGGT GGGGTGTAGT TCCTTTTAGT CAAAGGAAGA AAATGTTAGA GAAAGCTAGA
1561 GCAAAAAATA AAAAGCCTAA ATCCAGTGCT GGTATTTCTT CAATGCCGAA CATCACTGTT
1621 GGTACCCAGG TATGCTTGAG AAATAATCCT CTTTATCATG CTGGAGCAGT TGCATTTTCA
1681 ATTAGTGCTG GGATTCCTAA AGTTGGTGTC TTAATGGAGT CAGTTTGGAA TATGAATGAC
1741 AGCTGTAGAT TTCAACTTAG ATCTCCTGAA AGCTTGAAAA ACATGGAAAA AGCTAGCAAA
1801 ACTACTGAAG CTAAGCCTGA AAGTAAGCAG GAGCCAGTGA AAACAGAAAT GGGTCCTCCA
1861 CCATCTCCAG CATCCACGTG TAGTGATGCA TCCTCAATTG CCAGCAGTGC ATCAATGCCA
1921 TACAAACGAC GACGGTCAAC CCCTGCACCA AAAGAAGAGG AAAAGGTGAA TGAAGAGCAG
1981 TGGTCTCTTC GGGAAGTGGT TTTTGTGGAA GATGTCAAGA ATGTTCCTGT TGGCAAGGTG
2041 CTAAAAGTAG ATGGTGCCTA TGTTGCTGTA AAATTTCCAG GAACCTCCAG TAATACTAAC
2101 TGTCAGAACA GCTCTGGTCC AGATGCTGAC CCTTCTTCTC TCCTGCAGGA TTGTAGGTTA
2161 CTTAGAATTG ATGAATTGCA GGTTGTCAAA ACTGGTGGAA CACCGAAGGT TCCCGACTGT
2221 TTCCAAAGGA CTCCTAAAAA GCTTTGTATA CCTGAAAAAA CAGAAATATT AGCAGTGAAT
2281 GTAGATTCCA AAGGTGTTCA TGCTGTTCTG AAGACTGGAA ATTGGGTGCG ATACTGTATC
2341 TTTGATCTTG CTACAGGAAA AGCAGAACAG GAAAATAATT TTCCTACAAG CAGCATTGCT
2401 TTCCTTGGTC AGAATGAGAG GAATGTAGCC ATTTTCACTG CTGGACAGGA ATCTCCCATT
2461 ATTCTTCGAG ATGGAAATGG TACCATCTAC CCAATGGCCA AAGATTGCAT GGGAGGAATA
2521 AGGGATCCCG ATTGGCTGGA TCTTCCACCT ATTAGTAGTC TTGGAATGGG TGTGCATTCT
2581 TTAATAAATC TTCCTGCCAA TTCAACAATC AAAAAGAAAG CTGCTGTTAT CATCATGGCT
2641 GTAGAGAAAC AAACCTTAAT GCAACACATT CTGCGCTGTG ACTATGAGGC CTGTCGACAA
2701 TATCTAATGA ATCTTGAGCA ACGGTTTTTA GAGCAGAATC TACAGATGCT GCAGACATTC
2761 ATCAGCCACA GATGTGATGG AAATCGAAAT ATTTTGCATG CTTGTGTATC AGTTTGCTTT
2821 CCAACCAGCA ATAAAGAAAC TAAAGAAGAA GAGGAAGCGG AGCGTTCTGA AAGAAATACA
2881 TTTGCAGAAA GGCTTTCTGC TGTTGAGGCC ATTGCAAATG CAATATCAGT TGTTTCAAGT
2941 AATGGCCCAG GTAATCGGGC TGGATCATCA AGTAGCCGAA GTTTGAGATT ACGGGAAATG
3001 ATGAGACGTT CGTTGAGAGC AGCTGGTTTG GGTAGACATG AAGCTGGAGC TTCATCCAGT
3061 GACCACCAGG ATCCAGTTTC ACCCCCCATA GCTCCCCCTA GTTGGGTTCC TGACCCTCCT
3121 GCGATGGATC CTGATGGTGA CATTGATTTT ATCCTGGCCC CCGCTGTGGG ATCTCTTACC
3181 ACAGCAGCAA CCGGTACTGG TCAAGGACCA AGCACCTCCA CTATTCCAGG TCCTTCCACA
3241 GAGCCATCTG TAGTAGAATC CAAGGATCGA AAGGCGAATG CTCATTTTAT ATTGAAATTG
3301 TTATGTGACA GTGTGGTTCT CCAGCCCTAT CTACGAGAAC TTCTTTCTGC CAAGGATGCA
3361 AGAGGGATGA CCCCATTTAT GTCAGCTGTA AGTGGCCGAG CTTATCCTGC TGCAATTACC
3421 ATCTTAGAAA CTGCTCAGAA AATTGCAAAA GCTGAAATAT CCTCAAGTGA AAAAGAGGAA
```

FIGURE 3B CONTINUED

3481 GATGTATTCA TGGGAATGGT TTGCCCATCA GGTACCAACC CTGATGACTC TCCTTTATAT
3541 GTTTTATGTT GTAATGACAC TTGCAGTTTT ACATGGACTG GAGCAGAGCA CATTAACCAG
3601 GATATTTTTG AGTGTCGAAC TTGTGGCTTG CTGGAGTCAC TGTGTTGTTG TACGGAATGT
3661 GCAGGGGTTT GTCATAAAGG TCATGATTGG AAACTCAAAC GGACATCACC AACAGCCTAC
3721 TGTGACTGTT GGGAGAAATG TAAATGTAAA ACTCTTATTG CTGGACAGAA ATCTGCTCGT
3781 CTTGATCTAC TTTATCGCCT GCTCACTGCT ACTAATCTGG TTACTCTGCC AAACAGCAGG
3841 GGAGAGCACC TCTTACTATT CTTAGTACAG ACAGTCGCAA GGCAGACGGT GGAGCATTGT
3901 CAATACAGGC CACCTCGAAT CAGGGAAGAT CGTAACCGAA AAACAGCCAG TCCTGAAGAT
3961 TCAGATATGC CAGATCATGA TTTAGAGCCT CCAAGATTTG CCCAGCTTGC ATTGGAGCGT
4021 GTTCTACAGG ACTGGAATGC CTTGAAATCT ATGATTATGT TTGGGTCGCA GGAGAATAAA
4081 GACCCTCTTA GTGCCAGCAG TAGAATAGGC CATCTTTTGC CAGAAGAGCA AGTATACCTC
4141 AATCAGCAAA GTGGCACAAT TCGGCTGGAC TGTTTCACTC ATTGCCTTAT AGTTAAGTGT
4201 ACAGCAGATA TTTTGCTTTT AGATACTCTA CTAGGTACAC TAGTGAAAGA ACTCCAAAAC
4261 AAATATACAC CTGGACGTAG AGAAGAAGCT ATTGCTGTGA CAATGAGGTT TCTACGTTCA
4321 GTGGCAAGAG TTTTTGTTAT TCTGAGTGTG GAAATGGCTT CATCCAAAAA GAAAAACAAC
4381 TTTATTCCAC AGCCAATTGG AAAATGCAAG CGTGTATTCC AAGCATTGCT ACCTTACGCT
4441 GTGGAAGAAT TGTGCAACGT AGCAGAGTCA CTGATTGTTC CTGTCAGAAT GGGGATTGCT
4501 CGTCCAACTG CACCATTTAC CCTGGCTAGT ACTAGCATAG ATGCCATGCA GGGCAGTGAA
4561 GAATTATTTT CAGTGGAACC ACTGCCACCA CGACCATCAT CTGATCAGTC TAGCAGCTCC
4621 AGTCAGTCTC AGTCATCCTA CATCATCAGG AATCCACAGC AGAGGCGCAT CAGCCAGTCA
4681 CAGCCCGTTC GGGGCAGAGA TGAAGAACAG GATGATATTG TTTCAGCAGA TGTGGAAGAG
4741 GTTGAGGTGG TGGAGGGTGT GGCTGGAGAA GAGGATCATC ATGATGAACA GGAAGAACAC
4801 GGGGAAGAAA ATGCTGAGGC AGAGGGACAA CATGATGAGC ATGATGAAGA CGGGAGTGAT
4861 ATGGAGCTGG ACTTGTTAGC AGCAGCAGAA ACAGAAAGTG ATAGTGAAAG TAACCACAGC
4921 AACCAAGATA ATGCTAGTGG GCGCAGAAGC GTTGTCACTG CAGCAACTGC TGGTTCAGAA
4981 GCAGGAGCAA GCAGTGTTCC TGCCTTCTTT TCTGAAGATG ATTCTCAATC GAATGACTCA
5041 AGTGATTCTG ATAGCAGTAG TAGTCAGAGT GACGACATAG AACAGGAGAC CTTTATGCTT
5101 GATGAGCCAT TAGAAAGAAC CACAAATAGC TCCCATGCCA ATGGTGCTGC CCAAGCTCCC
5161 CGTTCAATGC AGTGGGCTGT CCGCAACACC CTGCATCAGC GAGCAGCCAG TACAGCCCCT
5221 TCCAGTACAT CTACACCAGC AGCAAGTTCA GCGGGTTTGA TTTATATTGA TCCTTCAAAC
5281 TTACGCCGGA GTGGTACCAT CAGTACAAGT GCTGCAGCTG CAGCAGCTGC TTTGGAAGCT
5341 AGCAACGCCA GCAGTTACCT AACATCTGCA AGCAGTTTAG CCAGGGCTTA CAGCATGTCA
5401 TTAGACAAAT CATCGGACTT GATGGGCCTT ATTCCTAAGT ATAATCACCT AGTATACTCT
5461 CAGATTCCAG CAGCTGTGAA ATTGACTTAC CAAGATGCAG TAAACTTACA GAACTATGTA
5521 GAAGAAAAGC TTATTCCCAC TTGGAACTGG ATGGTCAGTA TTATGGATTC TACTGAAGCT
5581 CAATTACGTT ATGGTTCTGC ATTAGCATCT GCTGGTGATC CTGGACATCC AAATCATCCT
5641 CTTCACGCTT CTCAGAATTC AGCGAGAAGA GAGAGGATGA CTGCGCGAGA AGAAGCTAGC
5701 TTACGAACAC TTGAAGGCAG ACGACGTGCC ACCTTGCTTA GCGCCCGTCA AGGAATGATG
5761 TCTGCACGAG GAGACTTCCT AAATTATGCT CTGTCTCTAA TGCGGTCTCA TAATGATGAG
5821 CATTCTGATG TTCTTCCAGT TTTGGATGTT TGCTCATTGA AGCATGTGGC ATATGTTTTT
5881 CAAGCACTTA TATACTGGAT TAAGGCAATG AATCAGCAGA CAACATTGGA TACACCTCAA
5941 CTAGAACGCA AAAGGACGCG AGAACTCTTG GAACTGGGTA TTGATAATGA AGATTCAGAA
6001 CATGAAAATG ATGATGACAC CAATCAAAGT GCTACTTTGA ATGATAAGGA TGATGACTCT
6061 CTTCCTGCAG AAACTGGCCA AAACCATCCA TTTTTCCGAC GTTCAGACTC CATGACATTC
6121 CTTGGGTGTA TACCCCCAAA TCCATTTGAA GTGCCTCTGG CTGAAGCCAT CCCCTTGGCT
6181 GATCAGCCAC ATCTGTTGCA GCCAAATGCT AGAAAGGAGG ATCTTTTTGG CCGTCCAAGT
6241 CAGGGTCTTT ATTCTTCATC TGCCAGTAGT GGGAAATGTT TAATGGAGGT TACAGTGGAT
6301 AGAAACTGCC TAGAGGTTCT TCCAACAAAA ATGTCTTATG CTGCCAATCT GAAAAATGTA
6361 ATGAACATGC AAAACCGGCA AAAAAAGAAG GGGAAGGAAC AGCCCGTGCT GCCAGAAGAA
6421 ACTGAGAGTT CAAAACCAGG GCCATCTGCT CATGATCTTG CTGCACAATT AAAAAGTAGC
6481 TTACTAGCAG AAATAGGACT TACTGAAAGT GAAGGGCCAC CTCTCACATC TTTCAGGCCA
6541 CAGTGTAGCT TTATGGGAAT GGTTCTTTCC CATGATATGC TGCTAGGACG TTGGCGCCTT
6601 TCTTTAGAAC TGTTCGGCAG GGTATTCATG GAAGATGTTG GAGCAGAACC TGGATCAATC
6661 CTAACTGAAT TGGGTGGTTT TGAGGTAAAA GAATCGAAAT TCCGCAGAGA AATGGAAAAA
6721 CTGAGAAACC AGCAGTCAAG AGATTTGTCA CTAGAGGTTG ATCGGGATCG AGATCTTCTC
6781 ATTCAGCAGA CTATGAGGCA GCTTAACAAT CACTTTGGTC GAAGATGTGC TACTATACCA

FIGURE 3B CONTINUED

```
6841 ATGGCTGTAC ACAGAGTAAA AGTCACATTT AAGGATGAGC CAGGAGAGGG CAGTGGTGTA
6901 GCACGAAGTT TTTATACAGC CATTGCACAA GCATTTTTAT CAAATGAAAA ATTGCCAAAT
6961 CTAGAGTGTA TCCAAAATGC CAACAAAGGC ACCCACACAA GTTTAATGCA GAGATTAAGG
7021 AACCGAGGAG AGAGAGACCG GGAAAGGGAG AGAGAAAGGG AAATGAGGAG GAGTAGTGGT
7081 TTGCGAGCAG GTTCTCGGAG GGACCGGGAT AGAGACTTTA GAAGACAGCT TTCCATCGAC
7141 ACTAGGCCCT TTAGACCAGC CTCTGAAGGG AATCCTAGCG ATGATCCTGA GCCTTTGCCA
7201 GCACATCGGC AGGCACTTGG AGAGAGGCTT TATCCTCGTG TACAAGCAAT GCAACCAGCA
7261 TTTGCAAGTA AAATCACTGG CATGTTGTTG GATTATCCCA GCTCAGCTGC TTCTCTTCTA
7321 GCAAGTGAGG ATTCTCTGAG AGCAAGAGTG GATGAGGCCA TGGAACTCAT TATTGCACAT
7381 GGACGGGAAA ATGGAGCTGA TAGTATCCTG GATCTTGGAT TAGTAGACTC CTCAGAAAAG
7441 GTACAGCAGG AAAACCGAAA GCGCCATGGC TCTAGTCGAA GTGTAGTAGA TATGGATTTA
7501 GATGATACAG ATGATGGTGA TGACAATGCC CCTTTGTTTT ACCAACCTGG GAAAAGAGGA
7561 TTTTATACTC CAAGGCCTGG CAAGAACACA GAAGCAAGGT TGAATTGTTT CAGAAACATT
7621 GGCAGGATTC TTGGACTATG TCTGTTACAG AATGAACTCT GTCCTATCAC ATTGAATAGA
7681 CATGTAATTA AAGTATTGCT TGGTAGAAAA GTCAATTGGC ATGATTTTGC TTTTTTTGAT
7741 CCTGTAATGT ATGAGAGTTT GCGGCAACTA ATCCTCGCGT CTCAGAGTTC AGATGCTGAT
7801 GCTGTTTTCT CAGCAATGGA TTTGGCATTT GCAATTGACC TGTGTAAAGA AGAAGGTGGA
7861 GGACAGGTTG AACTCATTCC TAATGGTGTA AAGAGACCAG TCACTCCACA GAATGTATAT
7921 GAGTATGTGC GGAAAGACGC AGAACACAGA ATGTTGGTAG TTGCAGAACA GCCCTTACAT
7981 GCAATGAGGA AAGGTCTACT AGATGTGCTT CCAAAAAATT CATTAGAAGA TTTAACGGCA
8041 GAAGATTTTA GGCTTTTGGT AAATGGCTGC GGTGAAGTCA ATGTGCAAAT GCTGATCAGT
8101 TTTACCTCTT TCAATGATGA ATCAGGAGAA AATGCTGAGA AGCTTCTGCA GTTCAAGCGT
8161 TGGTTCTGGT CAATAGTAGA GAAGATGAGC ATGACAGAAC GACAAGATCT TGTTTACTTT
8221 TGGACATCAA GCCCATCACT GCCAGCCAGT GAAGAAGGAT TCCAGCCTAT GCCCTCAATC
8281 ACAATAAGAC CACCAGATGA CCAACATCTT CCTACTGCAA ATACTTGCAT TTCTCGACTT
8341 TACGTCCCAC TCTATTCCTC TAAACAGATT CTCAAACAGA AATTGTTACT CGCCATTAAG
8401 ACCAAGAATT TTGGTTTTGT GTAGAGTATA AAAGTGTGT ATTGCTGTGT AATATTACTA
8461 GCAAATTTTG TAGATTTTTT TCCATTTGTC TAT
```

FIGURE 3C

1	MTSIHFVVHP	LPGTEDQLND	RLREVSEKLN	KYNLNSHPPL	NVLEQATIKQ	CVVGPNHAAF
61	LLEDGRVCRI	GFSVQPDRLE	LGKPDNNDGS	KLNSNSGAGR	TSRPGRTSDS	PWFLSGSETL
121	GRLAGNTLGS	RWSSGVGGSG	GGSSGRSSAG	ARDSRRQTRV	IRTGRDRGSG	LLGSQPQPVI
181	PASVIPEELI	SQAQVVLQGK	SRSVIIRELQ	RTNLDVNLAV	NNLLSRDDED	GDDGDDTASE
241	SYLAGEDLMS	LLDADIHSAH	PSVIIDADAM	FSEDISYFGY	PSFRRSSLSR	LGSSRVLLLP
301	LERDSELLRE	RESVLRLRER	RWLDGASFDN	ERGSTSKEGE	PNLDKKNTPV	QSPVSLGEDL
361	QWWPDKDGTK	FICIGALYSE	LLAVSSKGEL	YQWKWSESEP	YRNAQNPSLH	HPRATFLGLT
421	NEKIVLLSAN	SIRATVATEN	NKVATWVDET	LSSVASKLEH	TAQTYSELQG	ERIVSLHCCA
481	LYTCAQLENS	LYWWGVVPFS	QRKKMLEKAR	AKNKKPKSSA	GISSMPNITV	GTQVCLRNNP
541	LYHAGAVAFS	ISAGIPKVGV	LMESVWNMND	SCRFQLRSPE	SLKNMEKASK	TTEAKPESKQ
601	EPVKTEMGPP	PSPASTCSDA	SSIASSASMP	YKRRRSTPAP	KEEEKVNEEQ	WSLREVVFVE
661	DVKNVPVGKV	LKVDGAYVAV	KFPGTSSNTN	CQNSSGPDAD	PSSLLQDCRL	LRIDELQVVK
721	TGGTPKVPDC	FQRTPKKLCI	PEKTEILAVN	VDSKGVHAVL	KTGNWVRYCI	FDLATGKAEQ
781	ENNFPTSSIA	FLGQNERNVA	IFTAGQESPI	ILRDGNGTIY	PMAKDCMGGI	RDPDWLDLPP
841	ISSLGMGVHS	LINLPANSTI	KKKAAVIIMA	VEKQTLMQHI	LRCDYEACRQ	YLMNLEQAVV
901	LEQNLQMLQT	FISHRCDGNR	NILHACVSVC	FPTSNKETKE	EEEAERSERN	TFAERLSAVE
961	AIANAISVVS	SNGPGNRAGS	SSSRSLRLRE	MMRRSLRAAG	LGRHEAGASS	SDHQDPVSPP
1021	IAPPSWVPDP	PAMDPDGDID	FILAPAVGSL	TTAATGTGQG	PSTSTIPGPS	TEPSVVESKD
1081	RKANAHFILK	LLCDSVVLQP	YLRELLSAKD	ARGMTPFMSA	VSGRAYPAAI	TILETAQKIA
1141	KAEISSSEKE	EDVFMGMVCP	SGTNPDDSPL	YVLCCNDTCS	FTWTGAEHIN	QDIFECRTCG
1201	LLESLCCCTE	CARVCHKGHD	CKLKRTSPTA	YCDCWEKCKC	KTLIAGQKSA	RLDLLYRLLT
1261	ATNLVTLPNS	RGEHLLLFLV	QTVARQTVEH	CQYRPPRIRE	DRNRKTASPE	DSDMPDHDLE
1321	PPRFAQLALE	RVLQDWNALK	SMIMFGSQEN	KDPLSASSRI	GHLLPEEQVY	LNQQSGTIRL
1381	DCFTHCLIVK	CTADILLLDT	LLGTLVKELQ	NKYTPGRREE	AIAVTMRFLR	SVARVFVILS
1441	VEMASSKKKN	NFIPQPIGKC	KRVFQALLPY	AVEELCNVAE	SLIVPVRMGI	ARPTAPFTLA
1501	STSIDAMQGS	EELFSVEPLP	PRPSSDQSSS	SSQSQSSYII	RNPQQRRISQ	SQPVRGRDEE
1561	QDDIVSADVE	EVEVVEGVAG	EEDHHDEQEE	HGEENAEAEG	QHDEHDEDGS	DMELDLLAAA
1621	ETESDSESNH	SNQDNASGRR	SVVTAATAGS	EAGASSVPAF	FSEDDSQSND	SSDSDSSSSQ
1681	SDDIEQETFM	LDEPLERTTN	SSHANGAAQA	PRSMQWAVRN	TQHQRAASTA	PSSTSTPAAS
1741	SAGLIYIDPS	NLRRSGTIST	SAAAAAAALE	ASNASSYLTS	ASSLARAYSI	VIRQISDLMG
1801	LIPKYNHLVY	SQIPAAVKLT	YQDAVNLQNY	VEEKLIPTWN	WMVSIMDSTE	AQLRYGSALA
1861	SAGDPGHPNH	PLHASQNSAR	RERMTAREEA	SLRTLEGRRR	ATLLSARQGM	MSARGDFLNY
1921	ALSLMRSHND	EHSDVLPVLD	VCSLKHVAYV	FQALIYWIKA	MNQQTTLDTP	QLERKRTREL
1981	LELGIDNEDS	EHENDDDTNQ	SATLNDKDDD	SLPAETGQNH	PFFRRSDSMT	FLGCIPPNPF
2041	EVPLAEAIPL	ADQPHLLQPN	ARKEDLFGRP	SQGLYSSSAS	SGKCLMEVTV	DRNCLEVLPT
2101	KMSYAANLKN	VMNMQNRQKK	EGEEQPVLPE	ETESSKPGPS	AHDLAAQLKS	SLLAEIGLTE
2161	SEGPPLTSFR	PQCSFMGMVI	SHDMLLGRWR	LSLELFGRVF	MEDVGAEPGS	ILTELGGFEV
2221	KESKFRREME	KLRNQQSRDL	SLEVDRDRDL	LIQQTMRQLN	NHFGRRCATT	PMAVHRVKVT
2281	FKDEPGEGSG	VARSFYTAIA	QAFLSNEKLP	NLECIQNANK	GTHTSLMQRL	RNRGERDRER
2341	EREREMRRSS	GLRAGSRRDR	DRDFRRQLSI	DTRPFRPASE	GNPSDDPEPL	PAHRQALGER
2401	LYPRVQAMQP	AFASKITGML	LELSPAQLLL	LLASEDSLRA	RVDEAMELII	AHGRENGADS
2461	ILDLGLVDSS	EKVQQENRKR	HGSSRSVVDM	DLDDTDDGDD	NAPLFYQPGK	RGFYTPRPGK
2521	NTEARLNCFR	NIGRILGLCL	LQNELCPITL	NRHVIKVLLG	RKVNWHDFAF	FDPVMYESLR
2581	QLILASQSSD	ADAVFSAMDL	AFAIDLCKEE	GGGQVELIPN	GVNIPVTPQN	VYEYVRKYAE
2641	HRMLVVAEQP	LHAMRKGLLD	VLPKNSLEDL	TAEDFRLLVN	GCGEVNVQML	ISFTSFNDES
2701	GENAEKLLQF	KRWFWSIVEK	MSMTERQDLV	YFWTSSPSLP	ASEEGFQPMP	SITIRPPDDQ
2761	HLPTANTCIS	RLYVPLYSSK	QILKQKLLLA	IKTKNFGFV		

FIGURE 4

FIGURE 5A

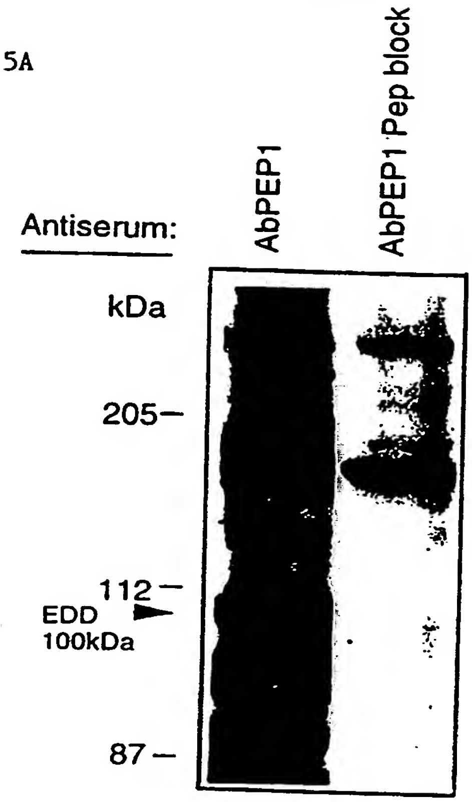
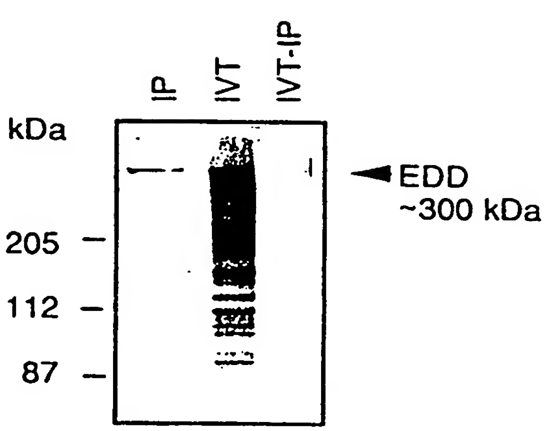


FIGURE 5 B



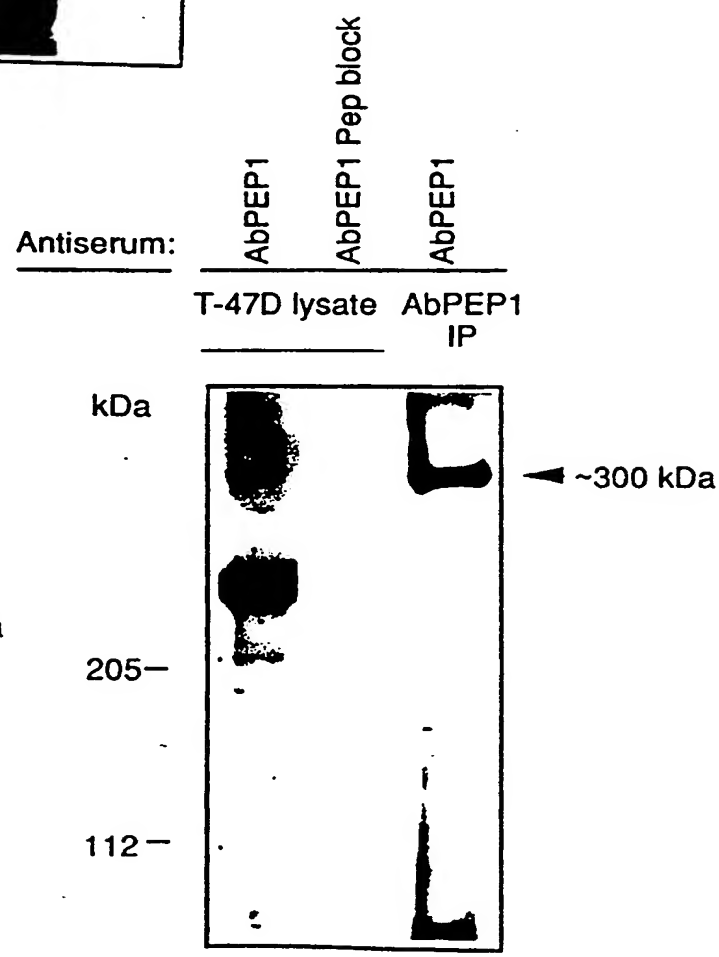


FIGURE 5C

FIGURE 6

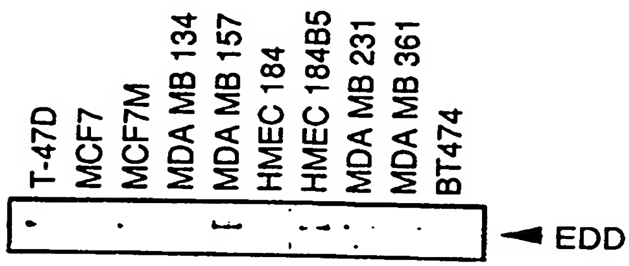


FIGURE 7

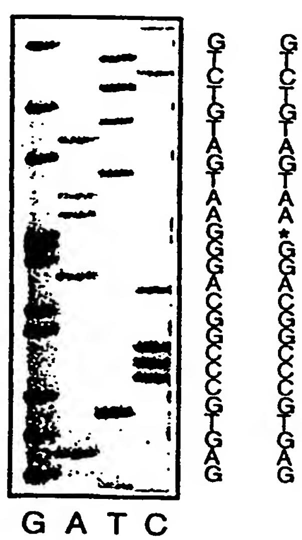


FIGURE 8A

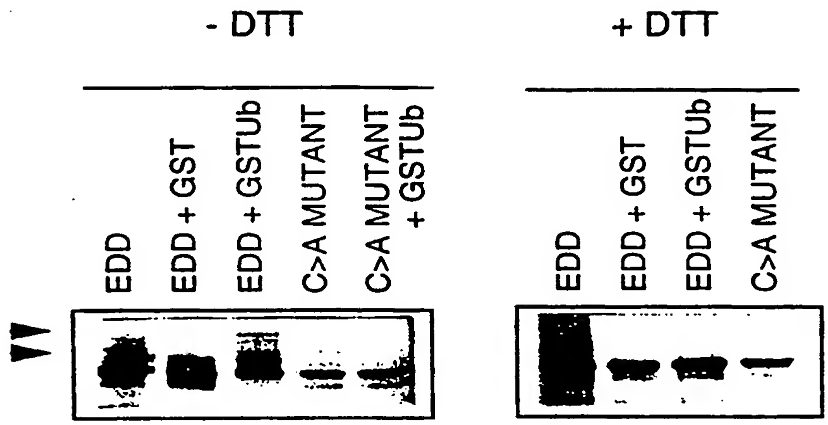


FIGURE 8B